



# ESCUELA SUPERIOR DE INGENIERÍA

## INGENIERÍA TÉCNICA EN INFORMÁTICA DE SISTEMAS

Detector y corrector automático de ediciones maliciosas en  
Wikipedia

Curso 2008-2009

Emilio José Rodríguez Posada

Cádiz, Enero de 2009



# ESCUELA SUPERIOR DE INGENIERÍA

## INGENIERÍA TÉCNICA EN INFORMÁTICA DE SISTEMAS

Detector y corrector automático de ediciones maliciosas en  
Wikipedia

Departamento: Lenguajes y Sistemas Informáticos.

Director del proyecto: Manuel Palomo Duarte.

Autor del proyecto: Emilio José Rodríguez Posada.

Cádiz, Enero de 2009

Fdo.: Emilio José Rodríguez Posada

Este documento ha sido liberado bajo Licencia GFDL 1.3 (GNU Free Documentation License). Se incluyen los términos de la licencia en inglés al final del mismo.

Copyright (c) 2009 Emilio José Rodríguez Posada.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

# Índice

<b>1. Introducción</b>	<b>9</b>
1.1. Objetivos . . . . .	9
1.2. Alcance . . . . .	9
1.3. Definiciones, acrónimos y abreviaturas . . . . .	10
1.4. Visión general . . . . .	15
<b>2. Desarrollo del calendario</b>	<b>17</b>
2.1. Iteraciones . . . . .	17
2.1.1. Iteración 1: Lectura de cambios recientes . . . . .	17
2.1.2. Iteración 2: Clasificación de usuarios . . . . .	17
2.1.3. Iteración 3: Distinción entre páginas nuevas o ediciones a páginas existentes . . . . .	18
2.1.4. Iteración 4: Otras acciones que no sean modificaciones de artículos . . . . .	18
2.1.5. Iteración 5: Análisis . . . . .	18
2.1.6. Iteración 6: Gestión de estadísticas . . . . .	19
2.1.7. Iteración 7: Contacto con los usuarios . . . . .	19
2.1.8. Iteración 8: Lista de expresiones regulares . . . . .	19
2.1.9. Iteración 9: Puntuación . . . . .	20
2.1.10. Iteración 10: Flexibilización de avisos . . . . .	20
2.1.11. Iteración 11: Autodiagnóstico . . . . .	21
2.1.12. Iteración 12: Lista de exclusiones . . . . .	21
2.2. Diagrama de Gantt . . . . .	21
2.3. Esfuerzo . . . . .	23
<b>3. Descripción general del proyecto</b>	<b>25</b>
3.1. Perspectiva del producto . . . . .	25
3.2. Funciones del producto . . . . .	25
3.3. Características de los usuarios . . . . .	28
3.4. Restricciones generales . . . . .	28
3.5. Suposiciones y dependencias . . . . .	29
3.6. Requisitos para futuras versiones . . . . .	29
<b>4. Desarrollo del proyecto</b>	<b>33</b>
4.1. Metodología de desarrollo . . . . .	33
4.2. Especificación de los requisitos del sistema . . . . .	33
4.3. Análisis y diseño del sistema . . . . .	34
4.3.1. avbot . . . . .	34

4.3.2.	avbotanalysis . . . . .	34
4.3.3.	avbotcomb . . . . .	35
4.3.4.	avbotglobals . . . . .	36
4.3.5.	avbotload . . . . .	36
4.3.6.	avbotmsg . . . . .	36
4.3.7.	avbotsave . . . . .	37
<b>5.</b>	<b>Codificación</b>	<b>39</b>
<b>6.</b>	<b>Pruebas y validación</b>	<b>42</b>
6.1.	Iteración 1: Lectura de cambios recientes . . . . .	43
6.2.	Iteración 2: Clasificación de usuarios . . . . .	43
6.3.	Iteración 3: Distinción entre páginas nuevas o ediciones a páginas existentes . . . . .	44
6.4.	Iteración 4: Otras acciones que no sean modificaciones de artículos . .	44
6.5.	Iteración 5: Análisis . . . . .	45
6.6.	Iteración 6: Gestión de estadísticas . . . . .	45
6.7.	Iteración 7: Contacto con los usuarios . . . . .	46
6.8.	Iteración 8: Lista de expresiones regulares . . . . .	47
6.9.	Iteración 9: Puntuación . . . . .	48
6.10.	Iteración 10: Flexibilización de avisos . . . . .	48
6.11.	Iteración 11: Autodiagnóstico . . . . .	48
6.12.	Iteración 12: Lista de exclusiones . . . . .	49
<b>7.</b>	<b>Conclusiones</b>	<b>51</b>
<b>8.</b>	<b>Manual de usuario</b>	<b>65</b>
8.1.	Ejecución . . . . .	65
8.2.	Lista de expresiones . . . . .	65
8.3.	Lista de mensajes . . . . .	66
8.4.	Lista de exclusiones . . . . .	67
8.5.	Estadísticas . . . . .	67
8.6.	Depuración . . . . .	67
8.7.	Para profundizar . . . . .	67
<b>9.</b>	<b>Manual de instalación</b>	<b>69</b>
<b>10.</b>	<b>Estadísticas</b>	<b>72</b>
10.1.	Artículos que han sufrido más intervenciones . . . . .	72

<b>11.¿Qué es Wikipedia?</b>	<b>75</b>
11.1. Historia . . . . .	76
11.2. Comunidad . . . . .	77
11.3. Control de calidad . . . . .	78
11.4. Críticas . . . . .	78
11.5. Proyectos hermanos . . . . .	79
11.6. Forks . . . . .	79
11.7. Wikipedia y los <i>bots</i> . . . . .	80
<b>12.Curso de edición de páginas MediaWiki</b>	<b>83</b>
12.1. Edición básica . . . . .	83
12.1.1. Formatos de texto . . . . .	83
12.1.2. Enlaces a otros artículos de Wikipedia . . . . .	83
12.1.3. Enlaces a otros artículos de Wikipedia en otros idiomas u otros proyectos de Wikimedia . . . . .	84
12.1.4. Enlaces a páginas web externas a Wikipedia . . . . .	84
12.1.5. Listas . . . . .	85
12.1.6. Secciones . . . . .	85
12.2. Imágenes . . . . .	85
12.3. Categorías . . . . .	86
12.4. <i>Interwikis</i> . . . . .	87
12.5. Tablas . . . . .	87
12.6. Plantillas . . . . .	88
12.7. Palabras mágicas . . . . .	89
12.8. Para profundizar . . . . .	89
12.7.1. Tiempo . . . . .	90
12.7.2. Estadísticas . . . . .	91
<b>13.GNU Free Documentation License</b>	<b>93</b>
13.1. APPLICABILITY AND DEFINITIONS . . . . .	93
13.2. VERBATIM COPYING . . . . .	95
13.3. COPYING IN QUANTITY . . . . .	95
13.4. MODIFICATIONS . . . . .	96
13.5. COMBINING DOCUMENTS . . . . .	98
13.6. COLLECTIONS OF DOCUMENTS . . . . .	98
13.7. AGGREGATION WITH INDEPENDENT WORKS . . . . .	98
13.8. TRANSLATION . . . . .	99
13.9. TERMINATION . . . . .	99
13.10FUTURE REVISIONS OF THIS LICENSE . . . . .	100

## Índice de figuras

1.	Diagrama de Gantt . . . . .	22
2.	Duración de las iteraciones y sus subtareas . . . . .	23
3.	Reversiones por clase de usuario durante 2008 . . . . .	27
4.	Estadísticas almacenadas por el programa . . . . .	46
5.	Resúmenes de edición en los que se detalla el motivo de la reversión .	46
6.	Ejemplo de informe generado automáticamente . . . . .	49
7.	Motivos de las reversiones de AVBOT y su porcentaje sobre el total .	72
8.	Página principal de Wikipedia en el año 2002 . . . . .	76
9.	Wikipedistas y miembros de Wikimedia Argentina . . . . .	78
10.	Ediciones a artículos por día durante 2008 . . . . .	81

## Índice de cuadros

1.	Artículos con más intervenciones . . . . .	73
2.	Formato de texto . . . . .	83
3.	Enlaces . . . . .	83
4.	Enlaces a proyectos hermanos . . . . .	84
5.	Enlaces externos . . . . .	84
6.	Listas . . . . .	85
7.	Secciones . . . . .	85
8.	Categorías . . . . .	86
9.	<i>Interwikis</i> . . . . .	87
10.	Palabras mágicas referentes a unidades temporales . . . . .	90
11.	Palabras mágicas referentes a datos estadísticos . . . . .	91



## 1. Introducción

### 1.1. Objetivos

Se pretende crear **un programa que proteja a Wikipedia en español**<sup>1</sup> (y por extensión a cualquier página web que utilice el software MediaWiki<sup>2</sup>) de usuarios maliciosos<sup>3</sup> que introduzcan modificaciones indeseadas. Para ello es necesario completar cada uno de los siguientes **objetivos**:

1. Lectura de los cambios que se producen en los artículos en tiempo real.
2. Análisis y valoración de cada uno de los cambios capturados en el «objetivo 1».
3. Tomar las acciones oportunas: actuar si el cambio es malicioso o dejarlo pasar si no supone peligro.
4. Mantener un *log* de las acciones realizadas para su depuración en caso de error.

Más adelante se hará una distinción entre modificaciones meramente maliciosas y ediciones de prueba. En cualquier caso, se trata de aportaciones que deben ser deshechas, pues no contribuyen a la mejora de Wikipedia.

### 1.2. Alcance

Se trata de un **proyecto innovador**, surge de la necesidad de controlar las ingentes cantidades de información que vuelcan miles de usuarios alrededor del mundo en un proyecto colaborativo como Wikipedia. Wikipedia pertenece a la denominada Web 2.0,<sup>4</sup> una nueva concepción de la World Wide Web<sup>5</sup> en la que los usuarios ya no son meros consumidores de datos que los webmasters cuelgan en Internet, sino que ellos mismos juegan el papel de productores/consumidores de esa información.

Esto conlleva una serie de ventajas y desventajas. Por un lado el potencial de producción crece exponencialmente,<sup>6</sup> ya que cualquier persona con un navegador web

---

<sup>1</sup>«Wikipedia» es un proyecto para crear una enciclopedia libre en multitud de idiomas.[1]

<sup>2</sup>«MediaWiki» es un software para la creación de *wikis*, con licencia GPL, programado en PHP y que utiliza bases de datos MySQL.[2][3]

<sup>3</sup>A estos «usuarios maliciosos» se les conoce en la jerga de Wikipedia como «vándalos», y a sus acciones como «vandalismos».[4]

<sup>4</sup>«Web 2.0» es un término acuñado por Tim O'Reilly y que hace referencia a una nueva generación de la web, caracterizada por una mayor participación de los usuarios en los contenidos.[5]

<sup>5</sup>La «World Wide Web», a grandes rasgos, es un sistema de documentos vinculados entre sí mediante hiperenlaces, que pueden ser consultados con cualquier navegador web.

<sup>6</sup>El crecimiento de Wikipedia en inglés desde octubre de 2002 hasta octubre de 2006 se asemejó a una curva exponencial. El número de artículos era doblado cada 346 días aproximadamente. Esto puede explicarse del siguiente modo: más contenido atrae más visitas, más visitas generan más ediciones, más ediciones crean más contenido.[6]

puede modificar los contenidos y mejorarlos. Por otro lado la cantidad de errores bienintencionados o de mala fe también crece.

Estas aportaciones no deseadas son, la mayoría de las veces, detectables fácilmente para un humano o un ordenador, ya que lo modificado resalta sobre lo demás por contener expresiones malsonantes o texto sin sentido. En otras ocasiones esta tarea se hace más compleja para un computador por no poder ser descrito mediante un patrón sencillo.

Existen muchas personas comprometidas con Wikipedia, y la mayoría de ellos controlan de manera altruista y voluntaria que nadie destruya los contenidos de la enciclopedia. Este esfuerzo podría dedicarse a la mejora o creación de nuevos artículos, en vez de a vigilar constantemente si los cambios producidos por usuarios nuevos o anónimos son malintencionados. Esto sería un gran beneficio para la comunidad. Para ello puede desarrollarse un programa que haga esta tarea, al menos con las ediciones claramente destructivas.

El producto de este proyecto tiene por nombre AVBOT. Es el acrónimo de «AntiVandal BOT» (en español: «BOT AntiVándalos»).[7] Este programa analizará las ediciones que se realicen en Wikipedia en español para buscar contribuciones maliciosas y las revertirá. **AVBOT no podrá sustituir nunca a una persona**, pues la capacidad de comprensión humana a la hora de leer un texto sobrepasa con creces la que pueda tener un programa de ordenador, aunque sí será de gran ayuda para revertir esos vandalismos repetitivos y monótonos.

### 1.3. Definiciones, acrónimos y abreviaturas

Para facilitar la comprensión del presente documento, se incluyen a continuación una serie de definiciones de términos, acrónimos y siglas que aparecen con frecuencia.

**Administrador** Persona que se dedica al mantenimiento de un sistema o red.[52] En Wikipedia hace referencia a aquellos usuarios que tienen ciertos *privilegios*<sup>8</sup> como borrar/restaurar páginas, bloquear/desbloquear usuarios, proteger/desproteger páginas, etc.[53] En Wikipedia en español se llaman «bibliotecarios».[54]

**API** Application Programming Interface. La API de **MediaWiki** proporciona acceso a los contenidos de las bases de datos.[56] Se sitúa en <http://es.wikipedia.org/w/api.php>.[12]

**Artículo** Página de Wikipedia con contenido enciclopédico. Su espacio de nombres tiene identificador «0».

---

<sup>7</sup>El término «AVBOT» surge de la fusión de varios conceptos. Primero la denominación de máquina como proceso automático (robot o «bot»), y «vándalo», palabra con la que se identifica a los usuarios que modifican Wikipedia con malas intenciones.

<sup>8</sup>Puede parecer que estos *poderes* dan ventaja a unos usuarios sobre otros. Nada más lejos de la realidad. Los bibliotecarios deben actuar acorde a unas normas y políticas votadas por la comunidad. Jimmy Wales llegó a decir «*ser administrador no es una gran cosa*».[55]

**ASCII** American Standard Code for Information Interchange

**AVBOT** Acrónimo de «AntiVandal BOT» («BOT AntiVándalos» en español).[57]

**Bibliotecario** Es el nombre con el que se conoce en Wikipedia en español a los **administradores**. [54]

**Blanqueo** Eliminación parcial o total del contenido de una página, lo que constituye una **edición maliciosa** o un acto de **vandalismo**. [58]

**Bloqueo** Suspensión temporal o indefinida a un usuario de su capacidad de modificar páginas. Los bloqueos sólo pueden realizarlos los **administradores** y siempre de manera justificada. [59] Existe un log público donde ver los últimos bloqueos realizados, [60] y los bloqueos vigentes. [61]

**Bot** Véase **robot**.

**Cambios recientes** Página especial de Wikipedia en la que puede observarse las modificaciones realizadas en cualquier parte de la enciclopedia en los últimos minutos. Para Wikipedia en español es <http://es.wikipedia.org/wiki/Special:RecentChanges>.

**Canal de IRC** Las comunicaciones dentro de un servidor de IRC se dividen en salas temáticas, a cada una de ellas se les llama canal. Por ejemplo, el canal de Wikipedia en español es **#wikipedia-es** dentro de la red [irc.freenode.org](http://irc.freenode.org).

**Citizendium** Proyecto para crear una enciclopedia libre enorme y fiable. [63]

**Commons** Véase **Wikimedia Commons**.

**Conflicto de edición** Error que se produce cuando dos o más personas modifican una misma página a la vez. Si uno de ellos guarda sus cambios, cuando el resto vaya a hacer lo mismo, recibirán un mensaje de advertencia, indicando que el contenido ha cambiado.

**CSS** Cascading Style Sheets, Hojas de estilo en cascada.

**Diff** Extracto que muestra las diferencias de contenido entre dos versiones distintas de una misma página.

**Discusión** Anexo a cualquier página del software MediaWiki desde donde se pueden discutir cambios en el contenido de dicha página.

**Enciclopedia libre** Compendio de conocimiento cuyo contenido se distribuye bajo una **licencia libre**. También se conoce con este nombre a un ***fork*** de Wikipedia en español. [62] Wikipedia no es el único proyecto que ha existido o existe con la intención de crear una enciclopedia libre. Con anterioridad a Wikipedia existió **Nupedia** [66] y recientemente se fundó **Citizendium**. [63]

**Edición** En el contexto de los *wikis* hace referencia a la modificación de una página.

**Edición maliciosa o destructiva** Véase **vandalismo**.

**Enlace interlingüístico** Vínculo que une a dos páginas sobre un mismo tema en distintos idiomas.

**Espacio de nombres** División que realiza el software MediaWiki para diferenciar distintos tipos de páginas. Los principales son: artículos, categorías, plantillas y páginas de usuario.

**Etiqueta** Nombre con el que se conoce a la **netiquette** en Wikipedia.[45]

**Expresión regular** Patrones con los que se pueden describir una gran variedad de cadenas en poco espacio.

**Fork** Software o proyecto que se separa en dos ramas de desarrollo distintas.<sup>9</sup>

**Fundación Wikimedia** «Organización sin ánimo de lucro dedicada a fomentar el crecimiento, desarrollo y distribución de contenido libre y políglota, y proveer la totalidad del contenido de estos proyectos wiki al público, libre de costes. La Fundación Wikimedia opera algunos de los más grandes proyectos de consulta en el mundo, incluyendo Wikipedia, el cuarto sitio más visitado de Internet.»<sup>10</sup>

**GNU** GNU's Not Unix

**GPL** GNU General Public License (Licencia Pública General GNU)

**GFDL** GNU Free Documentation License (Licencia de Documentación Libre de GNU)

**Historial** Conjunto formado por todas las versiones anteriores de una misma página, incluida la actual. Cada página mantiene su propio historial, y es utilizado frecuentemente para restaurar el contenido debido a vandalismos, desacuerdos en la redacción, etc.

**HTML** HyperText Markup Language[67]

**Interwiki** Véase **enlace interlingüístico**.

**IRC** Internet Relay Chat.[51] Freenode ([irc.freenode.org](http://irc.freenode.org)) es la red donde se sitúan los **canales** de Wikipedia y sus proyectos hermanos. Los canales de cambios recientes se encuentran en [irc.wikimedia.org](http://irc.wikimedia.org). [48, 49] Para acceder a ellos es necesario contar con un cliente de IRC, por ejemplo ChatZilla.

**Irclib** Librería que encapsula el protocolo **IRC**. [46]

**Licencia libre** Licencia que no establece unas restricciones férreas en cuanto a la libertad de uso, redistribución o modificación de una obra.

**Lista de seguimiento** Listado con las páginas que hemos decidido vigilar y sus cambios recientes.

**Los cinco pilares** Las cinco normas básicas de Wikipedia y en las que se sustentan el resto de **políticas**. (1) Wikipedia es una enciclopedia, (2) Wikipedia busca el **punto de vista neutral**, (3) Wikipedia es de contenido libre, (4) Wikipedia sigue

---

<sup>9</sup>Wikipedia en español tiene un *fork* llamado «Enciclopedia Libre».

<sup>10</sup>Extraído de su página oficial: «The Wikimedia Foundation, Inc. is a nonprofit charitable organization dedicated to encouraging the growth, development and distribution of free, multilingual content, and to providing the full content of these wiki-based projects to the public free of charge. The Wikimedia Foundation operates some of the largest collaboratively edited reference projects in the world, including Wikipedia, the fourth most visited website in the world.»[33]

unas normas de **etiqueta**, (5) Wikipedia no tiene normas firmes más allá de estos cinco pilares.[41]

**Manual de estilo** Conjunto de normas que ayudan a mantener un diseño uniforme entre los artículos de Wikipedia.[68]

**MediaWiki** Software para la creación de *wikis*, con licencia GPL, programado en PHP y que utiliza bases de datos MySQL.[2][3]

**MIT** Massachusetts Institute of Technology

**Namespace** Véase **espacio de nombres**.

**Netiquette** Conjunto de normas tácitas que regulan el comportamiento de una persona en Internet.[44]

**NPOV** Neutral Point Of View, véase **Punto de vista neutral**.

**Nupedia** Proyecto para crear una **enciclopedia libre** con artículos escritos por expertos y revisados por pares. Fue cerrado.[66]

**Política** Norma refrendada por la comunidad. Aunque Wikipedia no tiene normas firmes más allá de **los cinco pilares**, se recomienda seguir las políticas del proyecto, pues han sido elaboradas con un alto consenso y su no cumplimiento puede acarrear sanciones como bloqueos.

**Punto de vista neutral** Es uno de los **pilares de Wikipedia**. Según el texto de la **política** oficial de Wikipedia en español: «*El punto de vista neutral (PVN) establece que la enciclopedia debe contener hechos y que sus artículos deben ser escritos sin sesgos, presentando adecuadamente todos los puntos de vista existentes sobre tales hechos. [...] Esta política se malinterpreta con facilidad. No supone que sea posible escribir un artículo desde un único punto de vista objetivo no sesgado. Dice que debemos representar adecuadamente los diferentes puntos de vista y sin que el artículo afirme, implique o insinúe que alguno de ellos es el correcto. La neutralidad es mostrar todos los puntos de vista relevantes posibles tal y como son, para que cada lector adopte la opinión que prefiera*».[43] Jimmy Wales ha llegado a afirmar que es «*absoluto e innegociable*».[42]

**Python** Lenguaje de programación interpretado, publicado por primera vez por Guido van Rossum en 1991.[69]

**Pywikidiabot** Framework para desarrollar herramientas que hagan mantenimiento en Wikipedia y por extensión en cualquier sitio MediaWiki.[47]

**Redirección** Página que al ser visitada reenvía al navegador a otro destino. Se utiliza cuando un artículo puede tener varios títulos, de esta forma se evita duplicar el contenido.

**Regex** Véase **expresión regular**.

**Resumen de edición** Texto que puede adjuntarse a cada modificación de una página con el propósito de explicar en qué consisten los cambios realizados. Es útil cuando se consulta el **historial**. Se considera una buena práctica el rellenarlo.[71]

**Revertir** Restaurar el contenido de un artículo a su estado anterior.[77]

**Robot** Programa que realiza de manera automática tareas aburridas y tediosas.[72] Para ejecutar un bot en Wikipedia, se debe contar con el consenso de la comunidad, pues sus cambios son rápidos y pueden afectar a multitud de artículos en muy poco tiempo.[75, 73] Existen robots para poner interwikis,[74] crear artículos,[76] corregir faltas de ortografía, luchar contra el vandalismo y un sin fin de utilidades más.

**Sysop** Véase **administrador**.

**Trasladar** Acción de cambiar el título de una página. Para trasladar debe utilizarse la pestaña «mover» presente en cada artículo, de esta forma se conservan los historiales.[78]

**Unicode** Estándar con los principales sistemas de escritura del mundo.

**URL** Uniform Resource Locator

**Usuario anónimo** Usuario que no se ha registrado en Wikipedia y aparece identificado por su dirección IP.[79]

**Usuario malicioso** Véase **vándalo**.

**UTF** Unicode Transformation Format

**Vándalo** Usuario que comete **vandalismos**.

**Vandalismo** Modificación no deseada en la que se elimina parte de la información del artículo, se introducen palabras soeces, etc.[4]

**Web 2.0** Término acuñado por Tim O'Reilly y que hace referencia a una nueva generación de la web, caracterizada por una mayor participación de los usuarios en los contenidos.[5]

**Wiki** Sitio web que permite a sus visitantes modificar el contenido del mismo. Ward Cunningham, desarrollador del primer software wiki, llamado WikiWikiWeb, lo definió como «*the simplest online database that could possibly work*», [32] en español, «*la base de datos en línea más simple que puede funcionar*».

**Wikificar** Acción de adaptar y transformar un texto plano al formato **wiki** según las normas, **políticas** y **manual de estilo** del sitio.

**Wikimedia** Véase **Fundación Wikimedia**.

**Wikimedia Commons** Depósito de imágenes libres que pueden ser enlazadas desde los proyectos de la **Fundación Wikimedia** en cualquier idioma.

**Wikcionario** Proyecto para la creación de un diccionario libre.

**Wikibooks** Véase **Wikilibros**.

**Wikilibros** Proyecto para crear libros de texto educativos.

**Wikinews** Véase **Wikinoticias**.

**Wikinoticias** Fuente de noticias libre.

**Wikispecies** Directorio de especies libre.

**Wikipedia** Proyecto para la construcción de una **enciclopedia libre**. [1] Utiliza el software **MediaWiki**.

**Wiktionary** Véase Wikcionario.

**WWW** World Wide Web

### 1.4. Visión general

Tras esta «introducción» que da una visión global en la que se han presentado los «objetivos», «alcance del proyecto» y una serie de «definiciones» para facilitar la comprensión del texto, se incluyen con detalle el **resto de apartados** de la memoria.

Un aspecto importante del proyecto lo constituye el calendario, que contiene las **tareas que se han desempeñado**, su orden, precedencias y dependencias entre ellas. Esto se encuentra explicado en «Desarrollo del calendario» y ha sido generado usando la herramienta Planner.<sup>11</sup>

A continuación se presenta la «Descripción general del proyecto» que incluye la **perspectiva y funciones** del producto, características de los usuarios a los que va dirigido, restricciones generales, suposiciones, dependencias y requisitos para futuras versiones.

Luego se profundiza en el «Desarrollo del proyecto», con detalles de la **metodología** de desarrollo empleada, especificación de requisitos, análisis del sistema y diseño del sistema. En la sección de «Codificación» se comentan los aspectos más relevantes de esta etapa.

La parte de «Pruebas y validación» recoge datos de las más de **5000 horas de ejecución** que lleva acumuladas el programa y en el «resumen» y «conclusiones» se recopila lo más destacado del trabajo.

Son ineludibles los «manuales» de usuario e instalación, junto con una serie de anexos que ayudan a comprender mejor la función de este proyecto y el contexto en el que se desarrolla. Para cerrar el documento, se añadió el texto completo en inglés de la licencia GFDL 1.3.

---

<sup>11</sup> «Planner» es una herramienta libre para la gestión de proyectos.





## 2. Desarrollo del calendario

### 2.1. Iteraciones

Para la realización del presente proyecto se ha utilizado un modelo de **desarrollo iterativo incremental**. A continuación se detallan cada una de las etapas por las que ha ido pasando el software, y se observará como conforme se iba avanzando se añadían nuevas funcionalidades y se pulían las existentes.

#### 2.1.1. Iteración 1: Lectura de cambios recientes

El **primer paso** consistía en desarrollar la parte que permite **adquirir la lista de cambios recientes** de Wikipedia en español. Para ello se empezó creando el programa en un único módulo, importando las librerías que facilitan el acceso al **canal de IRC**, conectándonos al servidor y simplemente mostrando las líneas que se leían desde el canal.

**Esta versión no analizaba los datos** recibidos ni tomaba decisiones sobre éstos. Se ponía en comunicación con el **servidor de IRC**, y mostraba en pantalla un extracto de aquello que recibía: título del artículo modificado y nombre de usuario que ha realizado la modificación.

Hay que tener en cuenta que no sólo se recoge información acerca de ediciones, si no que en el canal se muestran *logs* de bloqueos, protecciones de páginas, traslados de artículos, etc. Estos elementos serán analizados más adelante.

#### 2.1.2. Iteración 2: Clasificación de usuarios

Un aspecto a tener en cuenta y básico a la hora de discernir entre ediciones bondadosas y maliciosas es quién las realiza. Un **administrador** de Wikipedia se supone que no cometerá vandalismos, mientras que un usuario nuevo (con escaso número de ediciones y del que no sabemos nada) es posible que cometa algún acto vandálico. Así pues **es necesario distinguir entre usuarios**: administradores, usuarios registrados y usuarios anónimos.

En esta versión se mostraba además de los dos datos de la «iteración 1», el número de ediciones del usuario. Un administrador suele tener más de 1000 ediciones en Wikipedia. Un usuario nuevo (registrado o no) parte de cero ediciones. Por lo tanto, averiguando este número podemos hacer una distinción precisa entre usuarios nóveles y veteranos, descartando las ediciones de aquellos usuarios comprometidos con el proyecto y que se supone no causarán daño deliberado a este. Para más seguridad se cargan dos listas públicas que contienen los *nicks* los administradores y de otros robots, evitándo el análisis de las contribuciones de estos usuarios.<sup>12</sup>

---

<sup>12</sup>Estas listas provienen de <http://es.wikipedia.org/wiki/Special:UserList>, desde donde

### 2.1.3. Iteración 3: Distinción entre páginas nuevas o ediciones a páginas existentes

Cuando alguien colabora con Wikipedia, puede que esté creando un artículo nuevo o que esté modificando uno ya existente. Dependiendo del caso en el que nos encontremos, tomaremos una decisión u otra. Quizás nos interese mejorar el estilo del artículo si es nuevo, pues muchos son creados por usuarios inexpertos y carecen del diseño apropiado.[68]

Es posible saber en qué caso nos encontramos fácilmente, pues el servidor de IRC nos informa si se trata de una página nueva poniendo un carácter especial, «N», junto al título del artículo.

De esta forma, capturando ese dato y llegados a este punto el programa ya sabe qué artículo ha sido modificado, quién ha realizado el cambio y si es una página nueva o no.

### 2.1.4. Iteración 4: Otras acciones que no sean modificaciones de artículos

Como se explicó anteriormente, el programa recoge más datos aparte de las meras ediciones a artículos. Se trata de información muy variada: bloqueos a usuarios, protecciones a páginas, altas de usuarios, traslados de artículos, borrados de páginas, etc. Todo esto se filtra y es analizado por un grupo de funciones distintas, para tomar decisiones apropiadas.

En esta etapa el programa ya sabe qué artículos son modificados, por quién, qué reputación tiene cada usuario, qué otras acciones se están realizando en Wikipedia en estos instantes (altas, borrados, bloqueos, ...), y va destinando cada una de ellas a un grupo de funciones que la analizarán, aunque todavía no se intervenga en el proceso.

### 2.1.5. Iteración 5: Análisis

Ahora es el momento de analizar todos esos datos que nos llegan. Sabemos qué artículo ha sido modificado y por quién. A partir de ahora el robot bajará el texto anterior a la modificación y la nueva versión del artículo para compararlas. Pueden darse varios casos:

1. El usuario **ha introducido texto nuevo**: vemos qué palabras contiene esa nueva aportación. ¿Contiene insultos o palabras soeces?[4] En caso afirmativo revertimos.
2. El usuario **ha retirado poco texto**: no hacemos nada.

---

se puede filtrar por tipo de usuario. Como se observa no sólo existen «bots» y «bibliotecarios», también hay «stewards», «burócratas», «check users» entre otros.

3. El usuario **ha retirado casi todo** el texto: puede tratarse de un blanqueo de contenido. Calculamos el porcentaje de texto retirado y si es abundante revertimos.

En esta versión el robot ya tomaba algunas decisiones basadas en unos algoritmos y una **lista de expresiones regulares** muy primitivos, revirtiendo aquellas ediciones que daban positivo.

### 2.1.6. Iteración 6: Gestión de estadísticas

Un aspecto importante y que no hay que descuidar son las **estadísticas**. Desde las primeras versiones del programa se introdujo una función que llevaba el conteo de ediciones totales, vandalismos, blanqueos, ediciones de prueba, etc. Todos estos datos eran impresos en una página especial de Wikipedia **cada 2, 12 y 24 horas**.<sup>[90]</sup> Dado que el robot se estaba ejecutando prácticamente las 24 horas del día, estos datos son continuos y han permitido hacer una serie de gráficas y extraer conclusiones realmente curiosas acerca del comportamiento y conductas de los vándalos.

### 2.1.7. Iteración 7: Contacto con los usuarios

Es costumbre en Wikipedia **avisar a los vándalos** en sus páginas de discusión cuando están realizando acciones dañinas, rogándoles que cesen en el empeño o serán bloqueados en breve.

Hasta ahora el programa se limitaba a deshacer las modificaciones dañinas en los artículos, pero llegados a este punto era necesario introducir unas modificaciones en las funciones de análisis, para que en el momento en el que una edición diera positivo, automáticamente enviar un mensaje de advertencia al vándalo: «si no te detienes, serás bloqueado por un administrador y no podrás seguir editando Wikipedia».

La explicación de porqué avisar al usuario es simple: muchos de ellos se detienen cuando se les informa que pueden ser bloqueados. De esta forma los administradores se ahorran de bloquear a muchos vándalos, pues estos desisten. Otros se rebelan y se vuelven más virulentos, peleando contra el robot, aunque afortunadamente estos son los menos.

### 2.1.8. Iteración 8: Lista de expresiones regulares

En un principio la **lista de expresiones regulares** que detecta las palabras soeces e insultos se cargaba desde un fichero de texto. En las primeras etapas del proyecto era un método suficiente, pero conforme se hacía más frecuente la necesidad de añadir nuevas expresiones a la lista el trabajo se hacía tedioso. Hay que tener en cuenta que el *script* se ejecuta desde un servidor al que hay que acceder mediante

protocolo FTP. Cada vez que se quería incorporar expresiones regulares nuevas había que modificar el fichero local, conectarse con un cliente de FTP, subir el fichero, detener el *bot*, ejecutarlo de nuevo y comprobar que todo había ido bien.

Como mejora se desarrolló unas funciones que capturaban la lista desde una página de la propia Wikipedia.<sup>13</sup> De esta forma, un usuario confiable puede modificar la lista de expresiones con su navegador web y el robot toma la lista actualizada al instante, generando un **informe** en el que detalla si las expresiones introducidas se han compilado correctamente o hay errores.

### 2.1.9. Iteración 9: Puntuación

Hasta ahora las decisiones sobre si revertir una edición o no se basaban en datos discretos. Si una expresión regular daba positivo, se revertía la edición. Esto dejaba bastante que desear, pues aunque el porcentaje de error no era alto, sí se podía reducir. La idea consistió en añadir un **sistema de puntuaciones** a las expresiones regulares. De este modo, para revertir una edición se debía alcanzar cierta cifra que se calculaba sumando todas las puntuaciones de todas las expresiones que daban positivo para una entrada concreta.

Además, no bastaba con llegar a cierto número como resultado, sino que en función de cuanto texto se hubiera añadido en la nueva edición, así de alto o bajo sería la cifra a alcanzar. Este sistema no es más que un **cálculo de densidad** de palabras soeces frente al total del texto insertado. Es posible que si un usuario añade 20 líneas de texto y entre ellas se encuentra una «palabra extraña», se trate de un edición realmente correcta. Pero si nos encontramos con un usuario que añade 20 caracteres a un artículo y en ese escaso texto ya hay una coincidencia con la lista de expresiones, muy posiblemente estaremos ante una edición maliciosa.

Por lo tanto mediante el juego de puntuaciones y el cálculo de densidades es posible amortiguar los falsos positivos, haciendo que el *bot* sea **más exacto** y revierta solamente aquello que es dañino para el proyecto.

### 2.1.10. Iteración 10: Flexibilización de avisos

En una iteración anterior se había incorporado un módulo para emitir avisos a los vándalos, indicándoles que su edición había sido revertida y que si continua será bloqueado. En principio era suficiente pero conforme el *bot* iba progresando se observó que algunas veces las ediciones si bien necesitaban ser revertidas, requerían realizar alguna puntualización al editor. Por ejemplo, puede tratarse de una edición de prueba en la que el usuario introdujo abundante texto aleatorio simplemente

---

<sup>13</sup>Las lista se encuentra en una subpágina que sólo pueden modificar los bibliotecarios y el controlador del *bot* en [http://es.wikipedia.org/wiki/Usuario:AVBOT/Lista\\_del\\_bien\\_y\\_del\\_mal](http://es.wikipedia.org/wiki/Usuario:AVBOT/Lista_del_bien_y_del_mal).

para aprender a manejar Wikipedia. En este caso el *bot* debe enviar un mensaje amigable en el que pida al recién llegado que realice las pruebas en la «Zona de pruebas» habilitada para estos casos. Por el contrario, si se trata de una edición meramente vandálica, el mensaje debe advertir que ese comportamiento no se tolera en Wikipedia.

Esta flexibilización de los avisos se consigue incorporando un parámetro junto a cada expresión regular, indicando si se trata de una «prueba» o de un «vandalismo». En principio con estas dos clasificaciones es suficiente, aunque no se descarta incorporar nuevas categorías.

#### **2.1.11. Iteración 11: Autodiagnóstico**

Cuando modificamos un artículo en Wikipedia, tenemos la posibilidad de dejar un pequeño comentario acerca de qué hemos cambiado, a modo de resumen. Cada vez que el *bot* revierte una edición, rellena ese resumen con información como el nombre del usuario que cometió el vandalismo, a qué versión se está restaurando, y un enlace a una página en la que se puede avisar si el robot está cometiendo errores.

Aunque en principio este sistema permitía hacer llegar al programador la mayoría de los informes de error que haya cometido el script, en esta etapa se desarrolló una función que se encarga de generar un informe automático cada vez que alguien deshace una edición del robot. De esta forma se facilita la depuración del programa.

#### **2.1.12. Iteración 12: Lista de exclusiones**

Llegados a este punto consideré oportuno crear una lista de exclusiones externa al script, ya que el sistema con el que se evitaba que el *bot* vigilara ciertas páginas no era fácilmente modificable (al igual que ocurrió en su momento con la lista de expresiones regulares). Creé una función que lee la lista de páginas que no se deben analizar. Esta información está almacenada en una zona especial de Wikipedia.

Con esta lista es posible gestionar de manera transparente aquellas páginas que no deben ser vigiladas por el programa, ya que suelen ser propicias a contener expresiones poco enciclopédicas, y que provocan falsos positivos. El robot toma la lista de exclusiones instantes después de que alguien la modifique, ya que la mantiene bajo su control.

De esta forma se reducen los errores de actuación del programa, y no interfiere en páginas que no le atañen explícitamente.

### **2.2. Diagrama de Gantt**

A continuación se incluye un diagrama de Gantt con el progreso de las iteraciones y tareas.

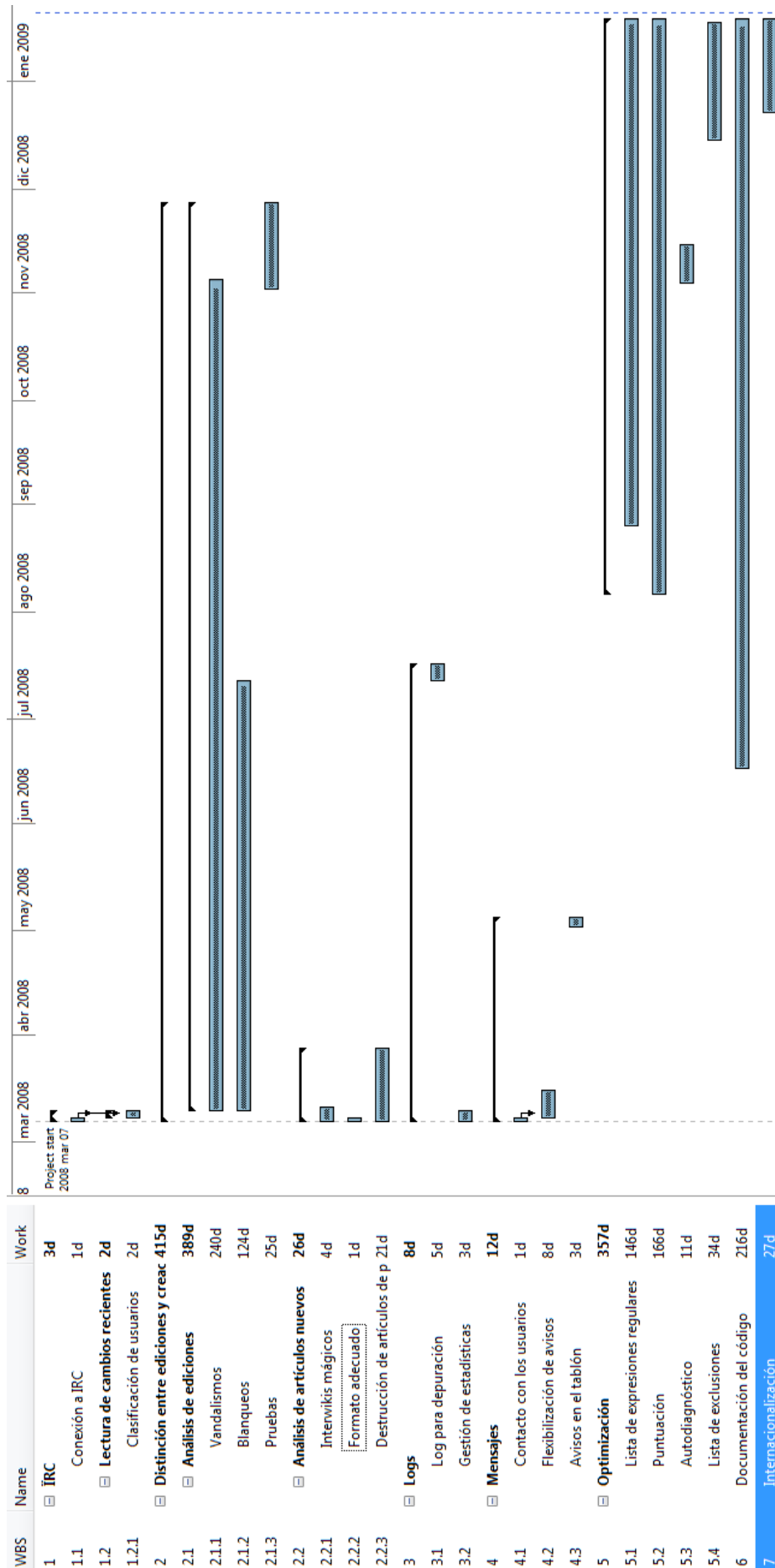


Figura 1: Diagrama de Gantt

Alguna de las tareas, como la lista de expresiones regulares o las puntuaciones no tienen fin. No se puede llegar a un punto en el que esté desarrollado al 100 %, pues son mejorables perpetuamente, de ahí que se les haya asignado un 99 % simbólico.

### 2.3. Esfuerzo

No es de extrañar que la mayor parte del **esfuerzo** se la haya llevado el **análisis de ediciones**, junto con la optimización de los sistemas de detección de vandalismos (lista de expresiones regulares y puntuaciones), ya que forman el núcleo del programa.

WBS	Name	Start	Finish	Work
1	IRC	mar 7	mar 9	3d
1.1	Conexión a IRC	mar 7	mar 7	1d
1.2	Lectura de cambios recientes	mar 7	mar 9	2d
1.2.1	Clasificación de usuarios	mar 7	mar 9	2d
2	Distinción entre ediciones y creaciones	mar 7	nov 26	415d
2.1	Análisis de ediciones	mar 10	nov 26	389d
2.1.1	Vandalismos	mar 10	nov 4	240d
2.1.2	Blanqueos	mar 10	jul 11	124d
2.1.3	Pruebas	nov 2	nov 26	25d
2.2	Análisis de artículos nuevos	mar 7	mar 27	26d
2.2.1	Interwikis mágicos	mar 7	mar 10	4d
2.2.2	Formato adecuado	mar 7	mar 7	1d
2.2.3	Destrucción de artículos de prueba	mar 7	mar 27	21d
3	Logs	mar 7	jul 16	8d
3.1	Log para depuración	jul 12	jul 16	5d
3.2	Gestión de estadísticas	mar 7	mar 9	3d
4	Mensajes	mar 7	may 4	12d
4.1	Contacto con los usuarios	mar 7	mar 7	1d
4.2	Flexibilización de avisos	mar 7	mar 16	8d
4.3	Avisos en el tablón	may 2	may 4	3d
5	Optimización	ago 6	ene 18	357d
5.1	Lista de expresiones regulares	ago 26	ene 18	146d
5.2	Puntuación	ago 6	ene 18	166d
5.3	Autodiagnóstico	nov 4	nov 14	11d
5.4	Lista de exclusiones	dic 15	ene 17	34d
6	Documentación del código	jun 17	ene 18	216d
7	Internacionalización	dic 23	ene 18	27d

Figura 2: Duración de las iteraciones y sus subtarefas





## 3. Descripción general del proyecto

### 3.1. Perspectiva del producto

El producto depende de unas librerías de las que importa una serie de funciones y clases que le permiten integrarse en un sistema mayor, Wikipedia. Es posible que éstas evolucionen y se requiera hacer modificaciones pues dado a que este es un mundo muy cambiante y en desarrollo perpetuo,[91] se hace necesario permanecer actualizado e incorporar estas mejoras al software.

Hay que tener en cuenta que este programa surge para cubrir una necesidad propia de los wikis, provocada por su carácter abierto. Tendrá sentido continuar el desarrollo de este proyecto mientras no se incorporen otras soluciones[16] que logren disminuir el vandalismo de manera más eficiente.

El programa funciona en modo consola. Se ha desestimado el desarrollo de una interfaz gráfica ya que ésta no ofrecería funcionalidad alguna, pues el software está diseñado para correr en segundo plano, sin intervención humana.

### 3.2. Funciones del producto

Dentro de la generalización que hacemos a la hora de hablar de «vandalismo», existe un abanico de posibilidades y un amplio espectro de distintos tipos de ediciones maliciosas. Escoger de entre un volumen de miles de ediciones perniciosas aquellas que comparten similitudes y extraer las características comunes nos permite analizarlas y **atajarlas en conjunto**, por lo que las posibilidades de éxito del robot crecen.

Tras analizar el comportamiento de los vándalos, se llega a la conclusión de que los principales tipos de vandalismos son los siguientes (acompañado de ejemplos inventados aunque para nada alejados de la realidad):

1. Introducción de **palabras soeces** o **insultos**:

- a) El vándalo llega al artículo titulado «Océano» y como no está satisfecho con el contenido, inserta la siguiente oración: «este artículo es una mierda».
- b) El vándalo está haciendo la tarea para su asignatura de Matemáticas, y tras leer el artículo sobre «Trigonometría» decide vengarse de un amigo de clase e incluye los siguiente: «Pedro es idiota».

2. Introducción de **texto aleatorio**:

- a) El usuario bien por malicia o por hacer una prueba de edición, elige un artículo e introduce texto sin sentido: «sidjfiogadsijgodsfgkadjngijsdifjod-jfigjdojg» o «aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa».

3. Realización de **pruebas de edición** de buena fe:

- a) El editor es novato y no sabe que existe una zona para hacer pruebas, así que elige un artículo al azar y pulsa varias veces en la barra de herramientas aleatoriamente: «”’[[Texto en negrita]]”’[[Archivo:Ejemplo.jpg]]».

4. **Blanqueo** de parte o totalidad del texto de un artículo. Pueden darse dos supuestos:

- a) El vándalo elimina un alto porcentaje del contenido del artículo.
- b) El vandalo elimina una parte o la totalidad de una sección en concreto de las varias que existen en el artículo.

5. Introducción de **datos personales** (fechas de cumpleaños, teléfonos, e-mails):

- a) En los artículos de días del año, como «23 de enero» o «4 de julio», una gran cantidad de las ediciones que se hacen son de usuarios que incluyen su fecha de nacimiento de la siguiente forma: «[[1988]] - Nace [[Luis Rodríguez]], el chico más inteligente de la Tierra.»
- b) Usuarios que dejan sus e-mails o teléfonos con distintos fines: «envíadme más información sobre loros a miemail@hotmail.com» o «mi celular es 666 XXX XXX quiero hacer amigos».

Se suele pensar que no es posible que haya tanta gente que acceda a Wikipedia, consulte los artículos y los destruya deliberadamente. En un mundo ideal esto no se produciría, pero los hechos demuestran que cada día, tan sólo en la edición en español de Wikipedia, el robot detecta entre 100 y 500 de estos vandalismos. Si contásemos los que revierten otros usuarios porque el programa no los captura serían más.

Aun así, cada día se realizan más de 20.000 ediciones en total, por lo que a *grosso modo* el porcentaje de ediciones maliciosas se sitúa con fluctuaciones (en función del día de la semana y de otros parámetros) entorno al 1-3 %. A pesar de lo catastrofista que pudiera parecer en un principio el proyecto es todo un éxito, de lo contrario habría colapsado por el vandalismo y esto no ha sucedido.

Además de la detección de este tipo de ediciones, el robot también realiza las siguientes funciones:

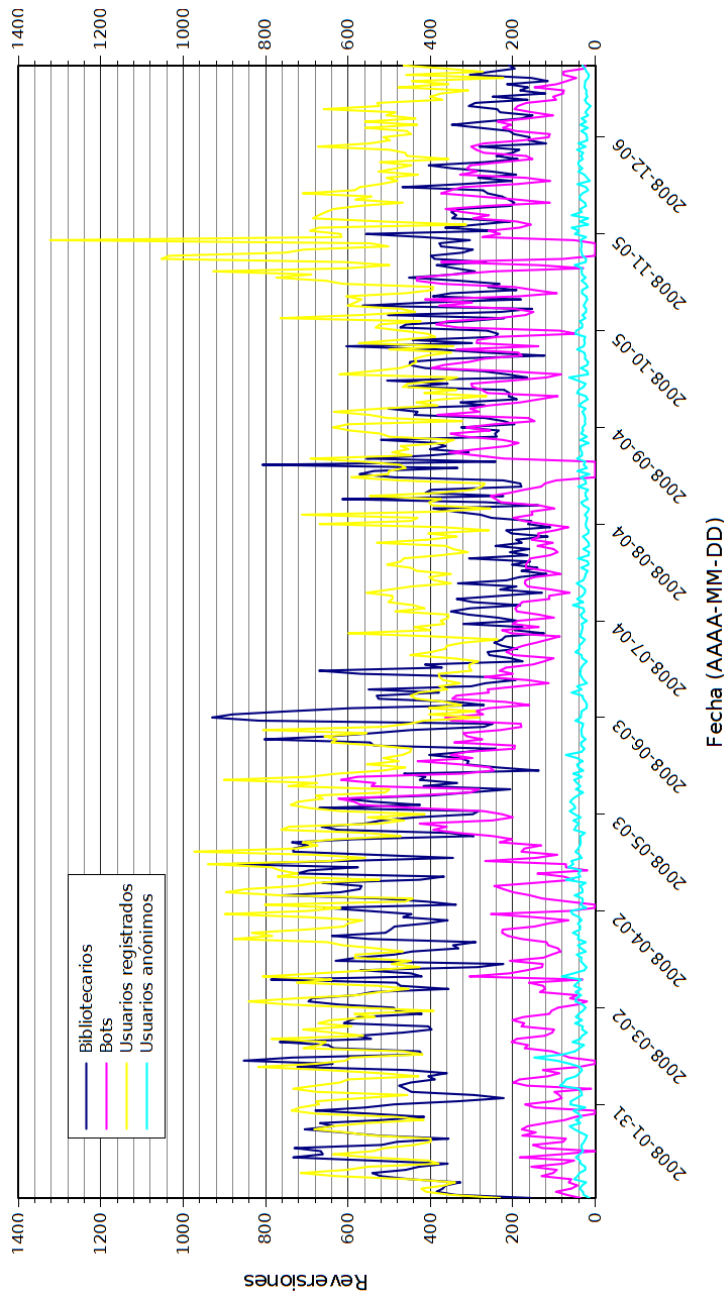


Figura 3: Reversiones por clase de usuario durante 2008

En la gráfica se observan las reversiones que se produjeron en Wikipedia en español cada día de 2008. La línea celeste representa aquellas que fueron realizadas por usuarios anónimos, y es la que menos se eleva de todas. Esto es así porque los anónimos no suelen dedicarse al mantenimiento de Wikipedia. La línea azul marino corresponde a los bibliotecarios y es muy similar a la rosa, correspondiente a AVBOT (desde marzo es el único robot que se dedica a revertir).

*¿No debería AVBOT hacer todo el trabajo? ¿Es que el robot no es eficiente?* Por desgracia no todas las ediciones inservibles son fácilmente identificables por un robot. Muchas de las reversiones que realizan los usuarios registrados (línea amarilla) o los bibliotecarios (línea azul marino) son reparaciones a ediciones que contienen frases fuera de contexto, no neutrales, vandalismos muy especializados, errores de novatos, etc, las cuales no pueden ser ponderadas por un programa. Tratar que el *bot* repare esos casos llevaría a muchos falsos positivos y sería contraproducente.

Con anterioridad a marzo existía un programa que hacía la misma función que AVBOT pero era menos eficiente (a penas alcanza las 200 reversiones/día). A partir de marzo y hasta fin de año se ve que AVBOT ha hecho casi 600 reversiones/día de media en los puntos más altos, aunque normalmente hace 300-400. Durante el periodo estival el número de vándalos descende, por lo que el robot tiene menos trabajo. A mediados de agosto y a finales de octubre el robot estuvo parado por mantenimiento de ahí que caiga hasta cero (se observa que la línea amarilla crece vertiginosamente para paliar la ausencia del robot). Las constantes subidas y bajadas (dientes de sierra) corresponden a la diferencia de actividad durante la semana (los miércoles son los máximos y los fines de semana los mínimos).

1. **Añadir *interwikis***<sup>14</sup> a artículos recién creados.
2. **Avisar a los vándalos** de que si no se detienen serán bloqueados<sup>15</sup> por un bibliotecario.
3. Avisar en la página «**Wikipedia:Vandalismo en curso**»<sup>16</sup> de aquellos vándalos reincidentes.
4. Hacer algunas **recomendaciones para usuarios novatos** que no se han clasificado como vándalos.
5. **Unificar** el nombre de las secciones «Véase también» y «Enlaces externos» de acuerdo al manual de estilo.

### 3.3. Características de los usuarios

A pesar de que el programa requiere una mínima intervención humana (más allá de los casos en los que sea necesario introducir nuevas expresiones regulares, o arreglar algún *bug*), el usuario ideal para administrar y ejecutar este software será aquel que disponga de una experiencia media o alta en wikis, preferiblemente en MediaWiki y en los proyectos de la Fundación Wikimedia (Wikipedia, Wikiquote, Wikcionario, etc), pues la relación entre AVBOT y éstos es estrecha.

También se recomienda tener nociones de HMTL y Python por si hubiera que realizar alguna modificación en el código. El conocimiento del funcionamiento de las plantillas en MediaWiki y su sintaxis es crucial. En un caso extremo podría ser necesario desarrollar algún *script* de apoyo o que reparara algún daño causado, aunque es poco probable.

### 3.4. Restricciones generales

La principal restricción del software es que «sólo»<sup>17</sup> funciona en MediaWiki. Para un rendimiento óptimo, el programa debería estar funcionando las 24 horas del día, 7 días de la semana, aunque esto necesita de un ordenador funcionando continuamente (la solución ideal es ejecutarlo desde un servidor). El sistema que ejecute este software debe tener acceso a IRC y una conexión a Internet suficiente para que no se produzcan problemas de *lag* y conflictos de edición.

---

<sup>14</sup>Un *interwiki* o «enlace interlingüístico» es un vínculo entre un artículo y su homólogo en otro idioma.

<sup>15</sup>La acción de bloquear a un usuario provoca que éste no pueda seguir modificando páginas de Wikipedia. Los bibliotecarios son los únicos que pueden bloquear a otros usuarios.

<sup>16</sup>Se trata de un apartado dentro de Wikipedia en el que se avisa a los bibliotecarios de que existe un usuario causando daño, para que actúen en consecuencia y lo bloqueen.

<sup>17</sup>Entrecomillado porque MediaWiki es uno de los sistemas wiki más extendidos, si no el que más.

### 3.5. Suposiciones y dependencias

El programa hace uso de una serie de librerías, por lo cual es necesario instalarlas para que el *script* funcione. Son, a saber: **pywikibiabot** e **irclib**, ambas libres, la primera con licencia del MIT, la segunda con licencia GPL.

Dado que el programa está desarrollado en Python, es ineludible tener instalado el intérprete de este lenguaje en nuestro sistema. Para las pruebas que se detallan en esta memoria, se utilizó la versión 2.5 del intérprete. Dado que el intérprete de Python es multiplataforma y está portado a sistemas operativos Windows y GNU/Linux, AVBOT es ejecutable en cualquiera de ambos sistemas.

Evidentemente se necesita tener registradas al menos dos cuentas de usuario en el proyecto, la que usará el robot para revertir y la del controlador. Será en sus subpáginas donde se crearán las listas de expresiones regulares y páginas excluidas, las estadísticas, etc.

Si el sitio web donde se va a ejecutar el programa requiere que la comunidad acepte el uso de este tipo de herramientas, se deberá pasar los cauces necesarios para ello.

### 3.6. Requisitos para futuras versiones

A pesar del funcionamiento favorable del programa, las posibles mejoras y nuevas funciones son varias, pues está sujeto a posibles cambios en el funcionamiento del software MediaWiki, así como intentar afrontar **nuevas formas de vandalismo** o recrudecimiento de las actuales.

A continuación se detalla algunas de las ideas para futuras características:

1. **Archivar** avisos de «Wikipedia:Vandalismo en curso» periódicamente. Dado al alto uso de esta página de avisos, es frecuente que alcance un gran tamaño, dificultando su consulta con el navegador web. La solución pasa por archivar los informes resueltos más antiguos y conservar los recientes.
2. **Controlar introducción masiva** de texto. AVBOT detecta blanqueos de texto, pero no la introducción de grandes cantidades de texto, a no ser que éste contenga palabras que concuerden con la lista de expresiones regulares.
3. Protección **anticumpleaños**. Evitar un frecuente tipo de vandalismo que consiste en introducir el noombre propio en las páginas de efemérides.
4. Control de **introducción de imágenes** de *hostings* externos (<http://imageshack.us>, <http://photobucket.com>, entre otras). No revertir estas ediciones, sólo avisar al usuario de que no funcionan este tipo de enlaces.

5. **Antiblanqueo de secciones.** A veces no se blanquea la totalidad del artículo, si no se ataca a una sola sección.
6. **Blanqueo de la parte final** de los artículos. Algunos vándalos suelen blanquear la última sección del artículo, la cual contiene las categorías e *interwikis*. Este tipo de vandalismo es un subconjunto del blanqueo de secciones.
7. **Revertir vandalismos recursivamente.** Si el robot no ha estado vigilando los cambios reciente continuamente, es posible que algún vandalismo se haya asentado en un artículo. De producirse un vandalismo posterior a este, el robot revertiría sólo el último, confiando en que la edición anterior es correcta.
8. Avisar de los **artículos más vandalizados** y sugerir que sean semiprotegidos. Cuando se producen una serie de vandalismos en cascada, es mejor optar por semiprotegerlo.
9. Controlar **imágenes chocantes**. Algunas imágenes anatómicas o médicas pueden resultar chocantes y son a menudo utilizadas por los vándalos con fines disruptivos. Cargando una lista con los nombres de estos archivos es posible evitar que sean insertados en páginas no relacionadas (páginas de usuario, en artículos con evidente falta de contexto, etc).
10. Detectar **v4nd4lim0s camuflados**. Algunos vándalos se divierten intentando superar la lógica del robot y empiezan a hacer pruebas para comprobar hasta dónde llega la capacidad del programa. Uno de los casos más frecuentes es camuflar el vandalismo cambiando caracteres, para saltarse las expresiones regulares.
11. Manejar las **funciones del bot mediante variables**. A pesar de que es posible modificar el comportamiento del *bot* en tiempo real, cambiando la lista de expresiones, la lista de páginas excluidas, etc, se podría mejorar la flexibilidad del programa haciendo una página de preferencias que se pudiera editar.
12. Evitar que pongan **{{destruir}}** en artículos con más de X ediciones. Un tipo de vandalismo posible aunque no frecuente es colocar la plantilla de destruir a un artículo que evidentemente debe conservarse. Por ejemplo el artículo sobre «España».
13. **Antitraslados**. Recientemente se ha venido observando un incremento de vandalismos basado en trasladar artículos muy visitados a títulos sin sentido. Controlando el log de traslados no sería difícil implementar una función que evitara estas ediciones.

14. **Rutas a ficheros locales.** En ocasiones, un usuario inexperto puede creer que escribiendo la ruta a un fichero en su ordenador, este se mostrará en Wikipedia. Es similar al caso de los hostings de imágenes. Un mensaje advirtiéndole del error sería suficiente.





## 4. Desarrollo del proyecto

### 4.1. Metodología de desarrollo

Para el desarrollo de este proyecto se ha empleado la **metodología *eXtreme Programming*** (programación extrema), ideada por Kent Beck. Ésta se basa en unos principios fundamentales, a saber: comunicación, simplicidad, retroalimentación y coraje.

- Se prefiere comenzar con una solución sencilla e ir ampliando posteriormente (simplicidad).
- Un código sencillo se explica por sí mismo (comunicación).
- Ciclos cortos permiten ver resultados con celeridad y centrarse en lo importante (retroalimentación).
- Confianza en que la programación por parejas es efectiva (coraje). Aunque este supuesto no se ha llevado a cabo en el presente proyecto.

El programa comenzó con una funcionalidad muy pequeña, como se ha visto en las iteraciones del calendario. El **desarrollo iterativo incremental** ha permitido una constante retroalimentación por parte de los resultados que se iban obteniendo y de las sugerencias que hacían los miembros de la comunidad.[96]

### 4.2. Especificación de los requisitos del sistema

Dando una visión global al proyecto, puede concluirse que los requisitos se agruparon del siguiente modo:

- **avbot** Módulo principal.
  - Dependencias: pywikipediabot e irclib
- **avbotanalysis** Módulo para detectar vandalismos, blanqueos, ediciones de prueba, y analizar páginas nuevas.
  - Dependencias: pywikipediabot.
- **avbotcomb** Módulo para funciones varias.
  - Dependencias: pywikipediabot.
- **avbotglobals** Módulo para variables compartidas.

- Dependencias: pywikipediabot.
- **avbotload** Módulo para carga de datos: lista de usuarios por número de ediciones, administradores, *bots*, lista de expresiones regulares y de páginas excluidas.
  - Dependencias: pywikipediabot.
- **avbotmsg** Módulo para enviar mensajes.
  - Dependencias: pywikipediabot.
- **avbotsave** Módulo para guardar datos: ediciones de usuarios y estadísticas.
  - Dependencias: pywikipediabot.

### 4.3. Análisis y diseño del sistema

A continuación se detallan las funciones y procedimientos que conforman cada módulo, con la descripción y lo que se espera de ellos.

#### 4.3.1. avbot

Módulo principal desde el que se llaman al resto de funciones.

- **avbot::main ( )** Crea un objeto BOT y lo lanza.

#### 4.3.2. avbotanalysis

Módulo para detectar vandalismos, blanqueos, ediciones de prueba y analizar páginas nuevas.

- **avbotanalysis::cleandiff( pageTitle, data )** Extrae el texto que ha sido insertado en la edición.
- **avbotanalysis::editAnalysis ( editData )** Lanzamiento de funciones de análisis previo descarte de páginas/ediciones que no deben ser analizadas.
- **avbotanalysis::improveNewArticle ( editData )** Intenta mejorar el artículo según unos consejos básicos del manual de estilo
- **avbotanalysis::isRubbish ( editData )** Analiza si se trata de un artículo nuevo inservible.

- **avbotanalysis::isSameVandalism** ( **regexlistold**, **regexlistnew** ) ¿Se trata del mismo vandalismo?
- **avbotanalysis::mustBeReverted** ( **editData**, **cleandata**, **userClass** ) ¿Debe ser revertida la edición? ¿Es vandalismo, prueba de edición o blanqueo?
- **avbotanalysis::newArticleAnalysis** ( **editData** ) Análisis de artículos nuevos
- **avbotanalysis::revertAllEditsByUser** ( **editData**, **userClass**, **regexlist** ) Revierte todas las ediciones de un usuario en un mismo artículo
- **avbotanalysis::sameOldid** ( **editData** ) ¿Es el mismo oldid?
- **avbotanalysis::updateStats** ( **type** ) Incrementa variables de estadísticas.
- **avbotanalysis::watch** ( **editData** ) ¿Debe ser analizada esta página?

#### 4.3.3. avbotcomb

Módulo para funciones varias que no encajan en ningún otro módulo.

- **avbotcomb::blockedUser** ( **blocker**, **blocked**, **castigo** ) Carga Vandalismo en curso y gestionar bloqueo.
- **avbotcomb::cleanLine** ( **line** ) Limpia una línea de IRC de basura.
- **avbotcomb::encodeLine** ( **line** ) Codifica una cadena en UTF-8 a poder ser.
- **avbotcomb::getParameters** ( ) Gestionar parámetros capturados de la consola.
- **avbotcomb::getTime** ( ) Coge la hora del sistema.
- **avbotcomb::getUserClass** ( **editData** ) Averigua el tipo de usuario del que se trata.
- **avbotcomb::magicInterwiki** ( **page**, **resumen**, **idioma** ) Buscar *interwikis* que pueden venirle bien al artículo.
- **avbotcomb::namespaceTranslator** ( **namespace** ) Carga espacios de nombres por idioma.
- **avbotcomb::resumeTranslator** ( **editData** ) Traductor de resúmenes de edición primitivo.

- **avbotcomb::semiprotect ( titulo, protector )** Pone la plantilla `{{semiprotegido}}` si no la tiene ya.
- **avbotcomb::vtee ( text, resumen )** Algunos cambios menores según el manual de estilo.

#### 4.3.4. avbotglobals

En este módulo no hay ninguna función, sólo variables compartidas.

#### 4.3.5. avbotload

Módulo para carga de datos: lista de usuarios por número de ediciones, administradores, *bots*, lista de expresiones regulares y de páginas excluidas.

- **avbotload::changedRegexpsList ( dic1, dic2 )** ¿Los dos diccionarios de expresiones regulares son distintos?
- **avbotload::loadBots ( )** Carga lista de *bots*.
- **avbotload::loadEdits ( )** Carga fichero con número de ediciones.
- **avbotload::loadExclusions ( )** Carga lista de páginas excluidas.
- **avbotload::loadMessages ( )** Carga preferencias sobre mensajes.
- **avbotload::loadRegexpList ( )** Carga lista de expresiones regulares.
- **avbotload::loadSysops ( )** Carga lista de administradores.
- **avbotload::loadUserEdits ( author )** Carga número de ediciones de un usuario en concreto.
- **avbotload::loadUsers ( type )** Captura lista de usuarios de Wikipedia según el tipo deseado.
- **avbotload::reloadRegexpList ( author, diff )** Recarga lista de expresiones regulares.

#### 4.3.6. avbotmsg

Módulo para enviar mensajes.

- **avbotmsg::msgBlock ( blocked, blocker )** Envía mensaje de bloqueo a un usuario.

- **avbotmsg::msgVandalismoEnCurso ( dic\_vand, author, userclass, blockedInEnglishWikipedia )** Gestiona la página de Vandaliso en curso.
- **avbotmsg::sendMessage ( author, wtitle, diff, n, tipo )** Envía mensajes de advertencia a un usuario.

### 4.3.7. avbotsave

Módulo para guardar datos: ediciones de usuarios y estadísticas.

- **avbotsave::saveEdits ( ediciones )** Guarda el número de ediciones de los usuarios en un fichero.
- **avbotsave::saveStats ( stats, hours, site )** Guarda las estadísticas en una página con motivos históricos.



## 5. Codificación

El *script* ha sido desarrollado en el **lenguaje de programación Python**,[34] el cual es interpretado, creado por Guido van Rossum en el año 1991. La versión del intérprete que se ha utilizado ha sido la 2.5.2. No se ha empleado ningún entorno de desarrollo integrado, en su lugar, se ha preferido usar un editor de textos avanzado con funciones específicas para programación llamado Notepad++,[40] que se distribuye con licencia GPL.

A la hora de consultar dudas sobre el lenguaje, se ha recurrido a la documentación oficial,[35] al tutorial *Python para todos* de Raúl González Duque,[36] y al documento Dive into Python.[37]

Algunos detalles interesantes de la codificación lo constituyen el uso de hilos y el precompilado de expresiones regulares.

En un principio el programa analizaba una edición tras otra, de forma secuencial. Este sistema era bastante lento, pues había que esperar a que se terminara de comprobar una edición para saltar a la siguiente y en Wikipedia en español el ritmo de modificaciones a veces superaba al ritmo de análisis. Para evitar esto se introdujo el uso de hilos, mediante la librería `thread`. Con esta solución, cada vez que llegan los datos de un cambio, se llama a la función de análisis. Si llegan varias modificaciones a la vez, no se espera a que termine una para empezar con la siguiente, ya que son independientes.

Otro aspecto relevante es el precompilado de expresiones. Cada vez que era analizada una edición, había que pasar la lista de expresiones (varios cientos de ellas) sobre un texto que en algunas ocasiones era bastante extenso. Este proceso requería los siguientes pasos:

1. Leer la expresión regular.
2. Compilar la expresión regular.
3. Pasar la expresión regular compilada al texto objetivo.
4. Volver al paso 1 con la siguiente expresión.

Se puede evitar la compilación continua de una misma expresión si guardamos el resultado de dicha compilación. Por ello, se implementó una mejora en el cargador de expresiones, que en vez de guardar solamente la expresión en cruda, guarda su compilación. La mejora fue drástica. De tener que compilar entorno a 600 expresiones por cada página analizada, lo que hacía que el proceso de análisis consumiera unos 10 segundos, se pasó a menos de una milésima de segundo por análisis. Definitivamente esta fue una mejora crucial en el programa.

Se han empleado globales, ya que de esta forma el código queda increíblemente más limpio, unido a que es un lenguaje de *scripting* y se hace uso intensivo de hilos.

Para finalizar, el código se ha organizado en módulos, cuyo contenido son funciones que guardan alguna relación entre sí. En un principio todo se encontraba en el mismo fichero, pero pronto quedó obsoleta esta organización.





## 6. Pruebas y validación

La finalidad primordial de una prueba de software es encontrar fallas en el sistema. Según las recomendaciones de G. J. Myers[97]:

1. Cada caso de prueba debe especificar el resultado esperado: En el caso que nos ocupa el resultado es simple, no error al revertir una edición. Revertir una edición legítima constituiría un error.
2. El programador debe evitar probar sus propios programas: El programa es puesto a prueba por todo aquel que edita Wikipedia, sin limitaciones.
3. Se debe analizar con detalle el resultado de cada prueba: Cada edición revertida satisfactoriamente es un éxito, cada reversión de una edición legítima es un error.
4. Se deben incluir entradas válidas e inválidas: El programador no controla lo que la gente va a introducir en Wikipedia. Es totalmente aleatorio.
5. Se debe:
  - a) Probar si el software no hace lo que debe: Un caso sería dejar pasar un vandalismo. La solución sería incorporar una expresión regular que lo capture.
  - b) Probar si el software hace lo que no debe: De nuevo, revertir una edición legítima. La solución pasa por afinar la lista de expresiones.
6. Los casos de prueba deben documentarse: Durante cada iteración del desarrollo se han ido analizando los datos anteriores e incorporando mejoras sobre ellos.
7. Hay que asumir que hay errores: Cada vez que alguien edita, está poniendo a prueba al programa.
8. Donde hay un defecto, hay otros: Cuando se ha encontrado un error, se ha revisado la zona afectada.
9. Las pruebas son una tarea creativa: Dada la naturaleza del proyecto, el cual recibe continuamente una serie de datos de entrada con cierta estructura pero cuyo contenido es completamente aleatorio (depende de lo que se le haya ocurrido teclear al humano, lo que constituye un «enfoque aleatorio»), hace que para la mayoría de los módulos no sea necesario que generemos nuestros casos de prueba (aunque algunos de ellos han sido analizados por separado, «prueba de unidad»), pues lo importante es ejecutar el programa *in the wild* («prueba

de integración») y ver si se comporta de manera adecuada. Por ello, y por el alto grado de éxito alcanzado, el programa ha acumulado más de 5000 horas de trabajo. De haber obtenido pobres resultados, la propia comunidad de Wikipedia en español hubiera desestimado el uso de esta herramienta y esto no ha sucedido («prueba de aceptación»).

Todas las clases, funciones y módulos que aparecen a continuación en negrita están detallados en la documentación.

## 6.1. Iteración 1: Lectura de cambios recientes

*¿Consigue conectarse al servidor de IRC? ¿Puede entrar al canal deseado? ¿Hay otros usuarios con el mismo nick? ¿Puede leer las líneas del canal? ¿Puede codificarlas a UTF-8? ¿Puede capturar el título y nick para luego mostrarlo en pantalla?*

`class BOT(SingleServerIRCBot):` Se creó una clase BOT que hereda de la clase `SingleServerIRCBOT` de la librería `irc.lib`. La conexión al servidor `irc.wikimedia.org` se realizó correctamente desde un primer momento, así como el acceso al canal `#es.wikipedia`. No se plantearon problemas de colisión de *nicks* con otros usuarios, pues se escogió uno novedoso (AVBOT). Aun así para evitar posibles errores de este tipo, se le añadió un número aleatorio de 4 cifras, quedando con el siguiente formato: AVBOT1234. Esto descarta cualquier posible fallo.

`BOT: :on_pubmsg:` En Wikipedia se utiliza la codificación UTF-8, por este motivo es importante trabajar internamente con ella, de modo que las líneas que se capturan del canal de IRC son convertidas a esta codificación. Python posee una serie de funciones para realizar conversiones entre cadenas (`unicode()` y `encode()`). Dado que al principio desconocía el uso de estas dos funciones, eran continuos los errores del programa cada vez que quería mostrar en pantalla una tilde o carácter especial, aunque funcionaba bien en los otros casos. Una vez superado este obstáculo, el tema de las conversiones entre codificaciones quedaría solucionado.

Para capturar el título y nick de usuario que realizó la modificación es necesario pasar una expresión regular a cada línea que se lee del canal de IRC. Esta expresión regular ha ido cambiando con el tiempo, unas veces porque el formato era modificado por los administradores del sitio, otras veces para optimizarla por voluntad propia.

## 6.2. Iteración 2: Clasificación de usuarios

*¿Puede descargar del servidor el número de ediciones de un usuario? ¿A partir de cuántas ediciones el usuario ya no es novato?*

`avbotload::loadUserEdits`: Hay varias formas de averiguar el número de ediciones de un usuario. Se escogió la más sencilla y directa, hacer una consulta a la API de Wikipedia y capturar el número. Esto se consigue descargando el contenido apuntado por cierta URL bien formateada (que contenga entre otros un parámetro que determine el nombre de usuario del que queremos obtener el dato). Hubo algún problema con los espacios de la URL (que se traducían a '%20'), y opté por convertirlos a guiones bajos '\_', por lo que el problema quedó solucionado. Lo mismo sucedía con aquellos nicks que contienen caracteres especiales. En estos casos hay que recodificarlos antes de incluirlos en la URL.

Decidir a partir de cuántas ediciones el usuario deja de ser novato no es una prueba de software, sino más bien un dato que se determina más o menos al azar, a través de las ejecuciones y de la experiencia propia en Wikipedia. Es un parámetro que se ha ido ajustando y refinando en sucesivas iteraciones.

### 6.3. Iteración 3: Distinción entre páginas nuevas o ediciones a páginas existentes

*¿Detecta cuando una página es nueva?*

`avbotpatterns::loadPatterns` y `BOT::on_pubmsg`: Cada línea que se captura posee la información sobre si la página es nueva o no, por ello, desde el momento en que la expresión regular que la analiza está bien formada, la comprobación de si es página nueva es trivial.

### 6.4. Iteración 4: Otras acciones que no sean modificaciones de artículos

*¿Captura bien todos los eventos y acciones? ¿Es capaz de mostrar esta información por pantalla?*

`avbotpatterns::loadPatterns` y `BOT::on_pubmsg`: En esta etapa el abanico de expresiones regulares se amplía. Es necesario crear una para cada acción: altas de usuarios, bloqueos, borrados de páginas, etc. En un primer momento esta fase fue propensa a errores, por la gran variedad de elementos a controlar, los cuales tenían distintos formatos. Para simplificarlo, se preprocesó cada línea para eliminar la información no deseada que resultaba un obstáculo (códigos de colores entre otros), hasta dejar la información cruda. Tras esto el análisis por expresiones regulares era mucho más fluido, y si algunas de las líneas aun era especial, era desviada a un log de errores para su posterior depuración a mano.

### 6.5. Iteración 5: Análisis

*¿Compara la versión antigua del texto con la nueva? ¿Es capaz de tomar decisiones en base a esta comparación? ¿Qué porcentaje de error tiene?*

`avbotanalysis::editAnalysis` y `avbotanalysis::mustBeReverted`: Este es el núcleo del programa y es la parte que se ha probado más exhaustivamente, de hecho es puesta a prueba continuamente cada vez que el robot toma una decisión y actúa.

Mediante la clasificación de usuarios se descartan a los veteranos y se analizan las ediciones de los recién llegados. En el supuesto de que el usuario haya introducido texto nuevo, éste debe pasar unos **filtros** compuestos de expresiones regulares que contienen palabras malsonantes. Aunque en toda la historia del programa los beneficios han sido mayores que los perjuicios (el robot no suele fallar más que un mínimo porcentaje de veces), en un primer estadio este porcentaje era mejorable. Cada vez que el *bot* revertía una edición, almacenaba en un *log* qué alarma había saltado y le había llevado a actuar (supuesto blanqueo de texto, supuesta palabra soez, etc), para su posterior análisis en caso de que hubiera sido una **falsa alarma**. Este método permite ir refinando la lista de expresiones regulares, eliminando aquellas que son propensas a provocar falsos positivos, incorporar nuevas, hacer mejoras en las existentes, o cambiar los «pesos» si se estima oportuno. También este «ensayo y error» continuo permitió ir ajustando el porcentaje de texto que hay que eliminar para que el robot actúe, y que se ha estimado en 6/7 del total.

### 6.6. Iteración 6: Gestión de estadísticas

*¿Lleva el conteo de ediciones, blanqueos, vandalismos y pruebas correctamente? ¿Los almacena?*

`avbotanalysis::updateStats` y `avbotsave::saveStats`: Las estadísticas se almacenan en un diccionario con tres claves (2, 12 y 24 horas) cuyo valor es a su vez un diccionario para guardar el conteo de «V» (vandalismos), «BL» (blanqueos), «P» (pruebas), «S» (contenido chocante), «B» (buenas), «M» (malas), «T» (total) y «D» (destruir). Cada vez que el robot lee una línea del canal de IRC que es detectada como edición, incrementa los contadores «T».<sup>18</sup> Cuando el programa llega a la conclusión de que una edición es vandalismo, blanqueo o prueba, incrementa los contadores «V», «BL» y «P», respectivamente. «M» y «B» es información derivada, y se calcula sumando «V», «BL» y «P» en el primer caso, y restando «T» menos

---

<sup>18</sup>Las líneas correspondientes a bloqueos, altas de usuarios, borrados de artículos, etc, no cuentan como edición.

«M» en el segundo.

Se escogió un periodo de dos horas al azar y llevando una cuenta manual de cada reversión y tipo (ya que cada vez que el *script* actúa, detalla en el resumen qué está haciendo). Se comparó con los datos que almacena Wikipedia, y fue satisfactorio, por lo que se concluyó que funciona correctamente.

Simbología: **(act)** = diferencia con la versión actual, **(prev)** = diferencia con la versión previa, **m** = edición menor

Comparar versiones seleccionadas

■	(act)	(prev)	13:35 8 ene 2009	AVBOT	(discusión   contribuciones)	m	(60 bytes)	(BOT - Actualizando estadísticas de las últimas 2 horas: V[7], BL[1], P[1], S[0], B[1338], M[9], T[1347], D[2])	(deshacer)
■	(act)	(prev)	11:35 8 ene 2009	AVBOT	(discusión   contribuciones)	m	(61 bytes)	(BOT - Actualizando estadísticas de las últimas 2 horas: V[9], BL[4], P[0], S[0], B[1064], M[13], T[1077], D[3])	(deshacer)
■	(act)	(prev)	09:35 8 ene 2009	AVBOT	(discusión   contribuciones)	m	(62 bytes)	(BOT - Actualizando estadísticas de las últimas 2 horas: V[10], BL[5], P[0], S[0], B[1302], M[15], T[1317], D[2])	(deshacer)

Figura 4: Estadísticas almacenadas por el programa

También se guardan cada 12 y 24 horas. Se resalta el periodo comprobado, que comienza a las 11:35 del 8 de enero de 2009 y finaliza a las 13:35. Se observa V[7], BL[1], P[1], S[0], B[1338], M[9], T[1347], D[2]. Extraído de <http://es.wikipedia.org/wiki/Usuario:AVBOT/Stats/2>.

■	13:47 8 ene 2009	(hist)	(dif)	m	Daniel Plaza	(BOT - Posible vandalismo de 213.97.96.59, revirtiendo hasta la edición 22462265 de Boto a Boto. ¿Hubo un error?)	(última modificación)
■	13:26 8 ene 2009	(hist)	(dif)	m	The devil wears Prada (película)	(BOT - Posible blanqueo de 190.55.231.59, revirtiendo hasta la edición 22778816 de 190.20.143.255. ¿Hubo un error?)	(última modificación)
■	13:10 8 ene 2009	(hist)	(dif)	m	Pila de combustible	(BOT - Posible vandalismo de 213.98.17.107, revirtiendo hasta la edición 22851595 de AVBOT. ¿Hubo un error?)	
■	13:02 8 ene 2009	(hist)	(dif)	m	Toyota RAV4	(BOT - Posible prueba de 190.11.93.186, revirtiendo hasta la edición 22196211 de Riad.Bot. ¿Hubo un error?)	(última modificación)
■	13:01 8 ene 2009	(hist)	(dif)	m	Estoma	(BOT - Posible vandalismo de 217.127.111.49, revirtiendo hasta la edición 23116593 de Saloca. ¿Hubo un error?)	
■	12:51 8 ene 2009	(hist)	(dif)	m	Estoma	(BOT - Posible vandalismo de 217.127.111.49, revirtiendo hasta la edición 23086334 de RoRo. ¿Hubo un error?)	
■	12:40 8 ene 2009	(hist)	(dif)	m	IEEE 802.3	(BOT - Posible vandalismo de 213.97.30.37, revirtiendo hasta la edición 21299006 de Thijs!bot. ¿Hubo un error?)	(última modificación)
■	12:25 8 ene 2009	(hist)	(dif)	m	Prometeo	(BOT - Posible vandalismo de 81.9.156.142, revirtiendo hasta la edición 22854267 de ArthurBot. ¿Hubo un error?)	
■	12:22 8 ene 2009	(hist)	(dif)	m	Sorteo Extraordinario de Navidad	(BOT - Posible vandalismo de 194.158.78.190, revirtiendo hasta la edición 22826966 de Kadellar. ¿Hubo un error?)	
■	11:44 8 ene 2009	(hist)	(dif)	m	Pau Donés	(BOT - Posible vandalismo de 85.86.36.15, revirtiendo hasta la edición 22759429 de AVBOT. ¿Hubo un error?)	(última modificación)
■	11:34 8 ene 2009	(hist)	(dif)	m	Energía eléctrica	(BOT - Posible vandalismo de 213.97.75.240, revirtiendo hasta la edición 22102070 de AVBOT. ¿Hubo un error?)	(última modificación)

Figura 5: Resúmenes de edición en los que se detalla el motivo de la reversión. Puede verse que coincide con la estadística generada. Hay 7 casos de vandalismo, 1 caso de blanqueo y 1 de prueba. Extraído de <http://es.wikipedia.org/wiki/Especial:Contributions/AVBOT>.

## 6.7. Iteración 7: Contacto con los usuarios

*¿Puede enviar mensajes a los usuarios cuyas ediciones han sido revertidas?*

`avbotmsg::sendMessage`: Cada vez que el *script* revierte una modificación envía un mensaje al autor de la edición que fue deshecha. Los mensajes contienen algunas instrucciones y advertencias, dependiendo del motivo que haya llevado a actuar al robot. Es trivial observar en la lista de contribuciones del programa que el envío se produce sin errores la mayoría de las veces. Aun así, se generaron unas estadísticas analizando los datos disponibles para noviembre y diciembre de 2008, resultando que se enviaron 11.078 mensajes para 10.973 reversiones. Supuestamente se enviaron 105 mensajes más de los debidos, pero la lectura correcta es que no pudo revertir 105 ediciones. ¿Por qué? Puede ser por varios motivos: algún usuario de los que se dedican a vigilar los cambios recientes se le adelantó, tuvo un conflicto de edición,<sup>19</sup> había *lag*,<sup>20</sup> etc. Son situaciones difícilmente controlables, que no tienen repercusión y que suponen menos del 1 % de los casos.

## 6.8. Iteración 8: Lista de expresiones regulares

*¿Puede leer y cargar la lista de expresiones regulares? ¿Puede detectar errores sin bloquearse? ¿Puede avisar de los errores?*

`avbotload::loadRegexpList` y `avbotload::reloadRegexpList`: Las **expresiones regulares** que hacen de filtro para detectar los casos de vandalismo y pruebas de edición se encuentran en una lista pública que pueden modificar solamente ciertos usuarios.<sup>21</sup> Cuando alguien modifica la lista, el programa la vuelve a leer para incorporar los cambios. Esta lectura se realiza línea a línea, **descartando los comentarios**<sup>22</sup> **y las expresiones regulares que no compilan**.

Para evitar que el *script* se cuelgue, la parte del programa que trata de compilar la expresión regular se encuentra dentro de una estructura «try-except», para capturar un posible error. En el caso de que así sea, añadirá información sobre éste suceso a una variable, cuyo contenido será impreso en una página anexa a la lista pública de expresiones, facilitando la depuración.<sup>23</sup>

---

<sup>19</sup>Un «conflicto de edición» se produce cuando dos usuarios están editando la misma página al mismo tiempo. Si el robot trata de enviar un mensaje a un usuario, y otra persona está enviando otro mensaje a ese mismo destinatario, sólo uno de ellos lo conseguirá. El otro recibirá un error.

<sup>20</sup>En ocasiones los servidores de Wikipedia están sobrecargados y no permiten editar las páginas, de ahí que algunas veces el programa no pueda guardar la reversión (el trabajo se pone en cola) pudiendo llegar a ser descartado o acabar en «conflicto de edición».

<sup>21</sup>En concreto los administradores del sitio y el controlador del *bot*. La lista se encuentra en [http://es.wikipedia.org/wiki/Usuario:AVBOT/Lista\\_del\\_bien\\_y\\_del\\_mal](http://es.wikipedia.org/wiki/Usuario:AVBOT/Lista_del_bien_y_del_mal).

<sup>22</sup>Para facilitar la comprensión de las expresiones presentes en la lista, se admiten comentarios precedidos por el carácter «#». Estas líneas son ignoradas por el programa.

<sup>23</sup>La salida de errores se situa en [http://es.wikipedia.org/wiki/Usuario\\_Discusi3n:AVBOT/Lista\\_del\\_bien\\_y\\_del\\_mal.css](http://es.wikipedia.org/wiki/Usuario_Discusi3n:AVBOT/Lista_del_bien_y_del_mal.css). Se puede ver un caso de éxito (<http://es.wikipedia.org/w/index.php?diff=23103924&oldid=23091129>) y un caso de error no deliberado (<http://es.wikipedia.org/w/index.php?diff=prev&oldid=19773708>).

## 6.9. Iteración 9: Puntuación

*¿Puede calcular la puntuación final de una edición y actuar en consecuencia?  
¿Es este método más eficaz y comete menos errores?*

`avbotanalysis::mustBeReverted`: El conteo de la puntuación se realiza mediante un bucle que recorre la lista de expresiones. Como cada expresión, tiene adjunta su puntuación, es trivial llevar un contador que sume (puntuaciones positivas o negativas, es indiferente). El *bot* revierte si se obtienen -5 puntos o menos. En caso de que sea 0 o mayor, no actúa. Si se encuentra entre -4 y -1, dependiendo de la cantidad de texto insertada, así se comportará. Estas cantidades se han fijado tras el análisis de los errores que se han ido produciendo. La implantación de esta característica consiguió reducir los errores hasta el 0,5 % actual. Definitivamente este sistema es mucho mejor y está más optimizado que el anterior.

## 6.10. Iteración 10: Flexibilización de avisos

*¿Detecta que tipo de edición ha realizado el usuario y envía el mensaje apropiado?*

`avbotanalysis::mustBeReverted` y `avbotmsg::sendMessage`: Por defecto se considera que una edición no tiene clase. Conforme se van pasando las expresiones regulares por el texto, en cuanto una da positivo, se almacena su clase. En función de la prioridad de cada clase, definida en «Mensajes.css», una prevalecerá sobre otra, independientemente del momento en que sea detectada.

El envío del mensaje apropiado está subordinado a la información que hayamos introducido en «Mensajes.css» y a la propia definición de la expresión regular, por lo que de producirse un error sería por alguna inconsistencia entre las clases de ambas páginas.

## 6.11. Iteración 11: Autodiagnóstico

*¿Detecta cuando un humano revierte al bot? ¿Es capaz de generar el informe con los datos básicos?*

Cada vez que se deshace una modificación, se deja tanto en el resumen de edición como en el aviso que se envía al usuario un enlace a una página donde se puede **informar de cualquier error** cometido por el *bot*. Para facilitar este trabajo de generación de informes y para evitar que algunos de ellos no lleguen a crearse por desidia, el programa detecta cuándo le revierten, emitiendo un informe automáti-



camente. Se almacenan en la página <http://es.wikipedia.org/wiki/Usuario:AVBOT/Errores/Automático> y permiten afinar aun más si cabe la depuración del código, indicando la fecha, el autor y el enlace al *diff*.

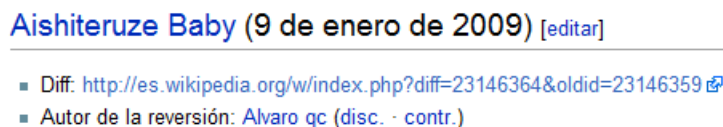


Figura 6: Ejemplo de informe generado automáticamente

## 6.12. Iteración 12: Lista de exclusiones

*¿Carga la lista de exclusiones y evita actuar sobre esas páginas? ¿Recarga la lista en cuanto es modificada?*

`avbotload::loadExclusions`: Al igual que las clases y funciones en un lenguaje de programación, las páginas de Wikipedia están clasificadas por espacios de nombres.<sup>24</sup>

El robot no debe vigilar cierto tipo de páginas, por ejemplo las de «discusión», ya que en ellas se utiliza lenguaje no enciclopédico y otras expresiones que provocarían falsos positivos. Evitar analizar grupos de páginas como éstos es fácil, pues se descartan según el identificador de su espacio de nombres. Pero hay ciertas páginas que aun perteneciendo a un espacio de nombres que debe ser vigilado, no sucede así con ellas. Para estos casos se implementó una «lista de exclusiones» (disponible en <http://es.wikipedia.org/wiki/Usuario:AVBOT/Exclusiones>) que permite insertar manualmente y una a una las páginas que deben ser ignoradas. Ejemplos: «Wikipedia:Zona de pruebas» o «Wikipedia:Libro de visitas».

---

<sup>24</sup>Entre ellos están: «Principal» (con identificador 0), «Usuario» (2), «Wikipedia» (4), «Archivo» (6), «MediaWiki» (8), «Plantilla» (10), «Ayuda» (12) y «Categoría» (14), entre otros personalizables. Los números impares se reservan para la página de discusión anexa correspondiente.



## 7. Conclusiones

La valoración del trabajo desarrollado es más que aceptable. Prueba de ello son las cifras alcanzadas:

- Más de 5000 horas de trabajo.
- Funcionamiento ininterrumpido comprobado durante más de 7 días (del 26 de noviembre al 2 de diciembre). Lo que dice mucho de su estabilidad.
- Más de 150 mil ediciones hasta la fecha. En torno a la mitad de ellas han sido reparaciones a páginas, el resto avisos a usuarios.
- Tasa de acierto del 99,5% según datos de diciembre de 2008. Primero fue mejorado con las puntuaciones, posteriormente con los contrapesos.
- Ningún bloqueo por malfuncionamiento.

Algunos de los usuarios de Wikipedia también han expresado su satisfacción con el trabajo del *bot*:

«Esto no es una sugerencia, sino una felicitación por los *interwikis mágicos* al crear un artículo. Muchas gracias.»

**Gerwoman**[95], +9000 ediciones

«Enhorabuena por la labor de ese robot y también felicito a su creador. Ha sido un gran paso adelante en la lucha contra vandalismos, haciéndoles más fácil la vida a los usuarios.»

**El Mexicano**[94], +1500 ediciones

«De verdad, que mil gracias por tanto trabajo que nos ahorra el *bot* y por tu preocupación de mantenerlo siempre trabajando del mejor modo posible.»

**Royfocker**[93], bibliotecario, +32000 ediciones

También pasó la autorización para poder ejecutarse con 19 votos a favor y 0 en contra.[92] Un signo de que la comunidad quiere que el *bot* continúe trabajando lo constituye la página de sugerencias,[96] desde donde se pueden pedir nuevas funcionalidades para el programa.

Sería de gran interés realizar algún estudio exhaustivo del histórico de ediciones revertidas desde los comienzos de Wikipedia. Esto proporcionaría una base

de conocimiento acerca de los más frecuentes tipos de vandalismo y lo mejor de todo, una ingente cantidad de posibles nuevas expresiones regulares y patrones que incorporar al programa.

También podría mejorarse la parte de detección de blanqueos basándose en ejemplos reales y teniendo en cuenta algún parámetro más además de la longitud total del texto y la del texto retirado.

En definitiva se trata de un producto cuyas cualidades son muy satisfactorias.



## Referencias

- [1] Wikipedia:About <http://en.wikipedia.org/w/index.php?oldid=261393469>  
Página «acerca de» en Wikipedia. Recoge sus fines, historia, datos sobre la comunidad, comparaciones con otras enciclopedias, datos técnicos, preguntas frecuentes, etc. 9, 14
- [2] MediaWiki.org <http://www.mediawiki.org/>  
Página principal del software MediaWiki en la que se puede encontrar, además de enlaces para su descarga en distintas versiones y documentación, extensiones que amplían sus funciones y lo hacen más potente. 9, 13
- [3] How does MediaWiki work? <http://www.mediawiki.org/w/index.php?oldid=228782>  
Página que recoge la licencia, lenguaje de programación, bases de datos y otros detalles técnicos del software MediaWiki. 9, 13
- [4] Wikipedia:Vandalism <http://en.wikipedia.org/w/index.php?oldid=262601744>  
Política de Wikipedia en inglés acerca de qué es el vandalismo, tipos, cómo luchar contra él, cómo no se debe reaccionar ante los vándalos y otras recomendaciones. 9, 14, 18
- [5] O'Reilly, Tim, «What Is Web 2.0», 2005, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>  
Documento sobre la llamada Web 2.0. Cómo se ha llegado hasta ella, qué es, comparaciones entre servicios de la Web 1.0 y la 2.0. 9, 14
- [6] Wikipedia:Modelling Wikipedia's growth <http://en.wikipedia.org/w/index.php?oldid=261063920>  
Página de Wikipedia en inglés en la que se estudia el crecimiento que ha tenido el proyecto durante los últimos años. 9
- [7] Fundación de Wikipedia en inglés, en alemán y en catalán <http://en.wikipedia.org/w/index.php?oldid=192353617>  
Estadísticas de las versiones de Wikipedia que existían en el año 2001. Se observa en qué mes comenzó a desarrollarse cada una. 76
- [8] Wikipedia en alemán en DVD <http://de.wikipedia.org/wiki/Wikipedia:Wikipedia-Distribution>  
Página de Wikipedia en alemán que detalla las características del DVD. 77

- [9] AFP/ELPAÍS.com, «La editorial Bertelsmann publicará en papel la Wikipedia alemana», 2008, [http://www.elpais.com/articulo/cultura/editorial/Bertelsmann/publicara/papel/Wikipedia/alemana/elpepucul/20080423elpepucul\\_11/Tes](http://www.elpais.com/articulo/cultura/editorial/Bertelsmann/publicara/papel/Wikipedia/alemana/elpepucul/20080423elpepucul_11/Tes)  
Noticia que recoge el proyecto de publicación de Wikipedia en alemán en papel. 77
- [10] Citas de Jimmy Wales en Wikiquote [http://en.wikiquote.org/wiki/Jimmy\\_Wales](http://en.wikiquote.org/wiki/Jimmy_Wales)  
Artículo de Wikicitas en el que se recogen algunas de las frases más destacadas pronunciadas por Jimmy Wales.
- [11] Donación a la Wikipedia para mejorar la usabilidad [http://wikimediafoundation.org/wiki/Press\\_releases/Wikipedia\\_to\\_become\\_more\\_user-friendly\\_for\\_new\\_volunteer\\_writers](http://wikimediafoundation.org/wiki/Press_releases/Wikipedia_to_become_more_user-friendly_for_new_volunteer_writers)  
Comunicado de prensa en el que se da cuenta de la donación de 890.000 dólares realizada por la *Stanton Foundation* para mejorar la usabilidad de Wikipedia y MediaWiki.
- [12] Barrapunto, «API para Wikipedia», 2008, <http://barrapunto.com/articles/08/04/26/2328245.shtml>  
Noticia que recoge la publicación de la API de Wikipedia. Realmente la API funciona para cualquier proyecto de la Fundación, ya que todos usan el software MediaWiki. 10
- [13] Barrapunto, «Actrices porno: la categoría más visitada en la Wikipedia en español», 2008, <http://barrapunto.com/articles/08/05/04/1131237.shtml>  
Noticia que relata el «descubrimiento» de que la categoría más consultada de Wikipedia en español es aquella que recoge las biografías de actrices porno. Se basa en un ranking realizado por el autor de esta memoria (<http://es.wikipedia.org/w/index.php?oldid=17100042>), analizando los *logs* de páginas vistas de Domas Mituzas (<http://dammit.lt/wikistats/>).
- [14] Barrapunto, «De dónde sale el tiempo para participar en Wikipedia», 2008, <http://barrapunto.com/articles/08/05/03/1717219.shtml>  
Noticia de un estudio sobre el tiempo libre y el uso que hace la gente de él.
- [15] Barrapunto, «El gobierno alemán dona 100.000 imágenes a la Wikipedia», 2008, <http://barrapunto.com/article.pl?sid=08/12/08/0114205>  
Noticia acerca de la donación de 100.000 imágenes a la Fundación Wikimedia, que serán alojadas en Wikimedia Commons. Realmente es una donación a la comunidad, pues las imágenes se han publicado con una licencia libre.

- [16] Wikipedia probará nuevo sistema de edición para disminuir el vandalismo [http://www.cooperativa.cl/prontus\\_notas/site/artic/20080718/pags/20080718104245.html](http://www.cooperativa.cl/prontus_notas/site/artic/20080718/pags/20080718104245.html)  
Noticia acerca de la extensión *FlaggedRevs* (<http://www.mediawiki.org/wiki/Extension:FlaggedRevs>) que permitiría controlar mejor los contenidos que se ingresan a Wikipedia evitando el vandalismo más eficazmente. 25
- [17] List of Wikipedias [http://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](http://meta.wikimedia.org/wiki/List_of_Wikipedias)  
Página de Meta-wiki (el sitio auxiliar de la Fundación Wikimedia para coordinar todos sus proyectos) en la que se mantiene una lista actualizada con el progreso en cuanto a número de artículos, usuarios, ediciones, imágenes locales, etc, de todas las Wikipedias en todos los idiomas. 75
- [18] wikipedia.org - Traffic Details from Alexa [http://www.alexa.com/data/details/traffic\\_details/wikipedia.org?range=5y&size=large&y=t](http://www.alexa.com/data/details/traffic_details/wikipedia.org?range=5y&size=large&y=t)  
Evolución de la posición de Wikipedia en el ranking de Alexa a lo largo del tiempo. 75
- [19] comScore Data Center <http://www.comscore.com/press/data.asp>  
Página que analiza los términos más populares de la red. 75
- [20] Alexa - TOP 500 [http://www.alexa.com/site/ds/top\\_sites?ts\\_mode=global&lang=none](http://www.alexa.com/site/ds/top_sites?ts_mode=global&lang=none)  
Ranking de Alexa de las 500 páginas web más visitadas del mundo. 75
- [21] Rodríguez Sánchez, Sergio, «La Wikipedia en la encrucijada», 2005, <http://www.elmundo.es/navegante/2005/12/19/esociedad/1134998202.html>  
Noticia de *El mundo* sobre la fiabilidad de Wikipedia. 75
- [22] BBC Mundo, «"Gratis y confiable", dice estudio», 2005, [http://news.bbc.co.uk/hi/spanish/science/newsid\\_4540000/4540140.stm](http://news.bbc.co.uk/hi/spanish/science/newsid_4540000/4540140.stm)  
Noticia que recoge el estudio de la revista *Nature* acerca de la fiabilidad de Wikipedia. 75, 79
- [23] Time's Person of the Year: You <http://www.time.com/time/magazine/article/0,9171,1569514,00.html> 75
- [24] Estadísticas de MediaWiki de Wikipedia en inglés <http://en.wikipedia.org/wiki/Special:Statistics>  
Estadísticas generadas automáticamente por el software MediaWiki que recogen el número de artículos, páginas, usuarios, usuarios activos en el último mes, entre otros datos acerca de Wikipedia en inglés. 77



- [25] Estadísticas de MediaWiki de Wikipedia en español <http://es.wikipedia.org/wiki/Special:Statistics>  
Versión de Wikipedia en español. 77, 80
- [26] EFE, «Vuelve la censura a la Wikipedia y a Blogspot tras dos meses de 'tregua'», 2006, <http://www.elmundo.es/navegante/2006/11/20/tecnologia/1164017826.html> 75
- [27] El País, «Wales se resiste a plegar la Wikipedia a la censura china», 2006, [http://www.elpais.com/articulo/internet/Wales/resiste/plegar/Wikipedia/censura/china/elpportec/20060831elpepnet\\_3/Tes](http://www.elpais.com/articulo/internet/Wales/resiste/plegar/Wikipedia/censura/china/elpportec/20060831elpepnet_3/Tes) 75
- [28] EFE, «La Wikipedia es accesible de nuevo en China tras un año de censura», 2006, <http://www.elmundo.es/navegante/2006/10/11/tecnologia/1160552915.html> 75
- [29] Wikipedia:Non-free content [http://en.wikipedia.org/wiki/Wikipedia:Non-free\\_content](http://en.wikipedia.org/wiki/Wikipedia:Non-free_content) 75
- [30] Wikipedia Founder Jimmy Wales Responds <http://interviews.slashdot.org/article.pl?sid=04/07/28/1351230> 75
- [31] ParserFunctions <http://www.mediawiki.org/wiki/Help:Extension:ParserFunctions>  
Documentación acerca de la extensión ParserFunctions. 91
- [32] What Is Wiki <http://www.wiki.org/wiki.cgi?WhatIsWiki>  
Definición de qué es un wiki. 14
- [33] <http://wikimediafoundation.org/wiki/Home>  
Página oficial de la Fundación Wikimedia. 12
- [34] Python.org <http://www.python.org/>  
Página oficial del lenguaje de programación Python. 39
- [35] Python Documentation <http://www.python.org/doc/>  
Documentación oficial del lenguaje de programación Python. 39
- [36] González Duque, Raúl, «Python para todos», <http://mundogeek.net/tutorial-python/>  
Tutorial de Python titulado «Python para todos». 39
- [37] Pilgrim, Mark, «Dive Into Python», 2004, <http://diveintopython.org/>  
Tutorial de Python. 39

- [38] Stallman, Richard, «Software libre para una sociedad libre», 2004, [http://www.gnu.org/philosophy/fsfs/free\\_software.es.pdf](http://www.gnu.org/philosophy/fsfs/free_software.es.pdf)
- [39] Raymond, Eric S., «La catedral y el bazar», <http://biblioweb.sindominio.net/telematica/catedral.html>  
Ensayo de Eric S. Raymond a favor del software libre.
- [40] Notepad++ <http://notepad-plus.sourceforge.net/>  
Página oficial del editor de textos avanzado Notepad++. 39
- [41] Wikipedia:Los cinco pilares <http://es.wikipedia.org/w/index.php?oldid=23014438>  
Página de Wikipedia que habla sobre «Los cinco pilares» fundamentales del proyecto. 13
- [42] [WikiEN-l] Articles about ourselves <http://lists.wikimedia.org/pipermail/wikien-l/2003-November/008096.html>  
Mensaje de Jimmy Wales en la lista de correo de Wikipedia en inglés afirmando que el «Punto de vista neutral» es «absoluto e innegociable». 13
- [43] Wikipedia:Punto de vista neutral <http://es.wikipedia.org/w/index.php?oldid=23093958>  
Política de Wikipedia sobre el «Punto de vista neutral». 13
- [44] Netiquette Guidelines <http://www.rfc-editor.org/rfc/rfc1855.txt> 13
- [45] Wikipedia:Etiqueta <http://es.wikipedia.org/w/index.php?oldid=23093989> 12
- [46] irclib <http://python-irclib.sourceforge.net/>  
Página oficial de irclib. 12
- [47] pywikipediabot <http://pywikipediabot.sourceforge.net/>  
Página oficial de pywikipediabot. 13
- [48] Meta:IRC <http://meta.wikimedia.org/wiki/IRC> 12
- [49] Meta:IRC channels [http://meta.wikimedia.org/wiki/IRC\\_channels](http://meta.wikimedia.org/wiki/IRC_channels) 12
- [50] Complementos para Firefox: ChatZilla <https://addons.mozilla.org/es-ES/firefox/addon/16>  
Página de la extensión de IRC ChatZilla para Firefox.
- [51] RFC 1459: Protocolo de Charla Basado en Internet <http://www.rfc-es.org/rfc/rfc1459-es.txt> 12

- [52] The Free On-line Dictionary of Computing <http://foldoc.org/index.cgi?system+administrator> 10
- [53] Wikipedia:Bibliotecarios <http://es.wikipedia.org/wiki/Wikipedia:Bibliotecarios>  
Página de Wikipedia sobre los bibliotecarios. 10
- [54] Wikipedia:Archivo de elecciones <http://es.wikipedia.org/w/index.php?oldid=21859994> 10, 11
- [55] [WikiEN-l] Sysop status <http://lists.wikimedia.org/pipermail/wikien-l/2003-February/001149.html>  
Mensaje de Jimmy Wales a la lista de correo de Wikipedia en inglés sobre el status de administrador. 10
- [56] Mediawiki:API <http://www.mediawiki.org/wiki/API>  
Documentación de la API de MediaWiki. 10
- [57] Página de usuario de AVBOT en Wikipedia en español <http://es.wikipedia.org/wiki/Usuario:AVBOT>  
Página de AVBOT dentro de Wikipedia en español. En ella se dan algunos datos, se listan las funciones que realiza, enlaces a las subpáginas de configuración, características futuras, gráficos, enlaces a la forja con el código fuente, etc. 11
- [58] Ejemplo de blanqueo <http://es.wikipedia.org/w/index.php?diff=prev&oldid=23305286> 11
- [59] Wikipedia:Blocking policy [http://en.wikipedia.org/wiki/Wikipedia:Blocking\\_policy](http://en.wikipedia.org/wiki/Wikipedia:Blocking_policy)  
Política sobre bloqueos en Wikipedia en inglés. 11
- [60] Log de bloqueos de Wikipedia en español <http://es.wikipedia.org/w/index.php?title=Special:Log&type=block>  
Registro con los últimos bloqueos realizados en Wikipedia en español. Se puede filtrar por meses y años, por autor del bloqueo o por usuario bloqueado. 11
- [61] Lista de bloqueos vigentes en Wikipedia en español <http://es.wikipedia.org/wiki/Special:BlockList>  
Registro con los bloqueos vigentes en Wikipedia en español. Es posible filtrar por nombre de usuario, IP, bloqueos de rangos y bloqueos temporales. 11
- [62] Enciclopedia Libre Universal en Español <http://enciclopedia.us.es/>  
Página principal de Enciclopedia Libre, el *fork* de Wikipedia en español. 11

- [63] Citizendium <http://www.citizendium.org/>  
Página principal de Citizendium. 11
- [64] Interpedia      FAQ      [http://groups.google.com/group/comp.infosystems.interpedia/browse\\_thread/thread/a0adc2203f87191c/e92420e978d2c3a0?lnk=st](http://groups.google.com/group/comp.infosystems.interpedia/browse_thread/thread/a0adc2203f87191c/e92420e978d2c3a0?lnk=st)  
Algunas preguntas y respuestas frecuentes acerca de Interpedia.
- [65] Wikipedia's heritage: vision, pragmatics, and happenstance <http://reagle.org/joseph/2005/historical/digital-works.html>
- [66] Nupedia en Internet Archive [http://web.archive.org/web/\\*/www.nupedia.com/](http://web.archive.org/web/*/www.nupedia.com/)  
Página principal de Nupedia en Internet Archive. 11, 13
- [67] HTML 4.01 Specification: What is HTML? <http://www.w3.org/TR/html401/intro/intro.html> 12
- [68] Wikipedia:Manual de estilo [http://es.wikipedia.org/wiki/Wikipedia:Manual\\_de\\_estilo](http://es.wikipedia.org/wiki/Wikipedia:Manual_de_estilo)  
Normas de estilo que se utilizan en Wikipedia en español para que los artículos sean uniformes. 13, 18, 78
- [69] Python History [http://svn.python.org/view/\\*checkout\\*/python/trunk/Misc/HISTORY](http://svn.python.org/view/*checkout*/python/trunk/Misc/HISTORY) 13
- [70] Mediawiki:Redirects <http://www.mediawiki.org/wiki/Help:Redirects>  
Documentación acerca de las redirecciones.
- [71] Meta:Edit sumary [http://meta.wikimedia.org/wiki/Help:Edit\\_summary](http://meta.wikimedia.org/wiki/Help:Edit_summary)  
Página de Meta-Wiki donde se explican los buenos usos del resumen de edición y algunos consejos. 13
- [72] Meta:Bot <http://meta.wikimedia.org/wiki/Bot>  
Página de Meta-Wiki acerca de los *bots*. 14, 80
- [73] Wikipedia:Bot      policy      [http://en.wikipedia.org/wiki/Wikipedia:Bot\\_policy](http://en.wikipedia.org/wiki/Wikipedia:Bot_policy)  
Política de Wikipedia en inglés acerca de los *bots*. Algunas definiciones, usos apropiados, proceso de aprobación... 14
- [74] Meta:Interwiki bot [http://meta.wikimedia.org/wiki/Interwiki\\_bot](http://meta.wikimedia.org/wiki/Interwiki_bot)  
Página de Meta-Wiki sobre los *bots* de *interwikis*. 14, 80

- [75] Meta:Bot policy [http://meta.wikimedia.org/wiki/Bot\\_policy](http://meta.wikimedia.org/wiki/Bot_policy)  
Página de Meta-Wiki con algunas recomendaciones generales sobre los *bots*. Autorización, usos inaceptables, intervalos de tiempo entre ediciones. 14
- [76] Meta:Botopedia <http://meta.wikimedia.org/wiki/Botopedia>  
Página de Meta-Wiki para coordinar la creación de artículos sobre pueblos del mundo a partir de datos estadísticos preformateados. 14
- [77] Meta:Reverting <http://meta.wikimedia.org/wiki/Help:Reverting>  
Página de Meta-Wiki que explica el proceso de revertir una edición. 14
- [78] Meta:Moving a page [http://meta.wikimedia.org/wiki/Help:Moving\\_a\\_page](http://meta.wikimedia.org/wiki/Help:Moving_a_page)  
Guía para hacer traslados correctamente. Trasladar sobre una redirección, trasladar a un título ocupado, deshacer un traslado, intercambiar los títulos de dos artículos. 14
- [79] Meta:Anonymous user [http://meta.wikimedia.org/wiki/Anonymous\\_user](http://meta.wikimedia.org/wiki/Anonymous_user)  
Descripción sobre usuarios anónimos en Meta-Wiki. 14
- [80] Bush, Vannevar, «As We May Think», 1945 <http://biblioweb.sindominio.net/pensamiento/vbush-es.html>  
Ensayo de Vannevar Bush de 1945 que se adelantó a su tiempo. Algunas de las descripciones coinciden con lo que hoy es la Web y Wikipedia.
- [81] The Free Universal Encyclopedia and Learning Resource <http://www.gnu.org/encyclopedia/free-encyclopedia.html>  
Ensayo de Richard Stallman sobre la creación de una enciclopedia libre.
- [82] Lista de correo de GNUPedia <http://lists.gnu.org/archive/html/info-gne/>
- [83] Wikipedia:Featured pictures [http://en.wikipedia.org/wiki/Wikipedia:Featured\\_pictures](http://en.wikipedia.org/wiki/Wikipedia:Featured_pictures)  
Página recopilatoria de Wikipedia en inglés con las imágenes más destacadas. 78
- [84] Commons:Featured pictures [http://commons.wikimedia.org/wiki/Commons:Featured\\_pictures](http://commons.wikimedia.org/wiki/Commons:Featured_pictures)  
Las mejores imágenes de Wikimedia Commons. 78
- [85] [Wikipedia-l] Wikipedia rocks! <http://lists.wikimedia.org/pipermail/wikipedia-l/2001-January/000000.html>

- Mensaje de Larry Sanger en el que presenta una página con los artículo que estaban mejor desarrolladosW por el año 2001. 78
- [86] Portal:Featured sounds [http://en.wikipedia.org/wiki/Portal:Featured\\_sounds](http://en.wikipedia.org/wiki/Portal:Featured_sounds)  
 Página de Wikipedia en inglés que muestra los mejores archivos de audio disponibles en el proyecto. 78
- [87] Wikipedia:Featured articles <http://en.wikipedia.org/w/index.php?oldid=264570472>  
 Página de Wikipedia en inglés que lista los artículos con calidad de «destacado». 78
- [88] Wikipedia:Good articles [http://en.wikipedia.org/wiki/Wikipedia:Good\\_articles](http://en.wikipedia.org/wiki/Wikipedia:Good_articles)  
 Página de Wikipedia en inglés que recoge aquellos artículos clasificados como «buenos». 78
- [89] Wikipedia:Stub <http://en.wikipedia.org/wiki/Wikipedia:Stub>  
 Guía sobre los «esbozos» en Wikipedia en inglés. Qué es, cómo marcar un artículo como esbozo, ... 78
- [90] Estadísticas de AVBOT cada 2 horas <http://es.wikipedia.org/w/index.php?title=Usuario:AVBOT/Stats/2&action=history>  
 Subpágina de AVBOT en la que se almacenan las estadísticas de las últimas 2 horas de ejecución. 19, 67
- [91] Wikipedia:Wikipedia is a work in progress [http://en.wikipedia.org/wiki/Wikipedia:Wikipedia\\_is\\_a\\_work\\_in\\_progress](http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_a_work_in_progress)  
 Ensayo en el que se detalla la naturaleza de continuo desarrollo de Wikipedia. 25
- [92] Votación de AVBOT [http://es.wikipedia.org/wiki/Wikipedia:Bot/Autorizaciones/Archivo\\_2008](http://es.wikipedia.org/wiki/Wikipedia:Bot/Autorizaciones/Archivo_2008)  
 Votación sobre si se permite ejecutar AVBOT en Wikipedia en español. 19 votos a favor, 0 en contra. 51
- [93] Agradecimiento de Royfocker <http://es.wikipedia.org/w/index.php?diff=23225154&oldid=23225141> 51
- [94] Agradecimiento de El Mexicano <http://es.wikipedia.org/w/index.php?diff=22222041&oldid=22122516> 51

- [95] Agradecimiento de Gerwoman <http://es.wikipedia.org/w/index.php?diff=21260716&oldid=21224702> 51
- [96] Sugerencias para AVBOT <http://es.wikipedia.org/wiki/Usuario:AVBOT/Sugerencias> 33, 51
- [97] Myers, G. J., «The Art of Software Testing», 1979 42





## 8. Manual de usuario

Se incluye un **manual de usuario** con las instrucciones necesarias para poner en ejecución el programa, así como algunos consejos para sacarle el mejor partido.

### 8.1. Ejecución

Es posible ejecutar el robot en cualquier idioma y en cualquiera de los proyectos de la Fundación (Wikipedia, Wikcionario, Wikicitas, Wikiversidad...) aunque por defecto actúe en la versión en español de Wikipedia si se omite los parámetros. Para ello debemos modificar el fichero «user-config.py» de nuestra instalación de pywikibot, tal y como se explica en la sección «Manual de instalación». A parte deberemos usar los siguientes parámetros:

- lang** Indica el idioma del proyecto. Suele ser el subdominio asociado. Para Wikipedia en francés sería «fr», de <http://fr.wikipedia.org>. Por defecto «es».
- family** Indica el proyecto. Puede ser Wikipedia, Wiktionary, Wikiquote, Wikiversity, etc. Por defecto «wikipedia».

A continuación se detallan algunos ejemplos de ejecución:

`avbot.py` Se ejecuta por defecto sobre Wikipedia en español.

`avbot.py -lang:fr` Se ejecuta sobre Wikipedia en francés.

`avbot.py -family:wikiquote` Se ejecutaría sobre Wikiquote en español.

`avbot.py -lang:de -family:wiktionary` Funciona en Wiktionary en alemán.

`avbot.py -lang:it -family:wikiquote` Analizaría Wikiquote en italiano.

Además de los parámetros «lang» y «family», existen los siguientes:

- botnick** Indica el nombre de usuario del robot. Por defecto «AVBOT».
- ownernick** Indica el nombre de usuario del controlador del *bot*. Por defecto «Owner».
- newbie** Indica el número de ediciones que hay que tener para ser considerado novato. Por defecto «25».
- statsdelay** Cada cuantos segundos se mostrará por pantalla información estadística. Por defecto «60».
- network** Red de IRC en la que se muestran los cambios recientes. Por defecto «irc.wikimedia.org».
- channel** Canal dentro de la red de IRC anterior en la que se publican los cambios. Por defecto se genera a partir de «lang» y «family».

### 8.2. Lista de expresiones

La **lista de expresiones regulares** disponible en «Lista del bien y del mal.css» es completamente modificable, además de que **los cambios son aplicados al in-**

**stante.** Esto es útil por si estamos bajo un ataque que se basa en algún patrón fácilmente identificable. Se permiten comentarios precedidos por el carácter «#». El formato que debe tener cada línea de la lista es el siguiente:

**CLASE;;EXPRESIÓNREGULAR;;PUNTUACIÓN;;**

Donde cada parte representa lo siguiente:

1. «CLASE»: Determina qué mensaje dejará el robot en la discusión de la persona que editó el artículo. El aviso de «Vandalismo» puede ser más brusco que el de «Prueba». Si alguien escribe «hola» en un artículo, debería recibir el aviso de prueba para ser orientado a la «Zona de pruebas». La clase puede ser:
  - a) «V» (Vandalismo)
  - b) «B» (Blanqueo)
  - c) «P» (Prueba)
  - d) «C» (Contrapeso)

Es posible añadir más mensajes o modificar los existentes en la subpágina «Mensajes.css».

2. «EXPRESIÓNREGULAR»: No es *case sensitive*. Hay que proteger los caracteres especiales con una barra invertida «\».
3. «PUNTUACIÓN»: Un número. Las posibilidades son las siguientes:
  - a) Un número positivo (ejemplos: 1, 2, 10, 100). Para premiar.
  - b) Un número negativo (ejemplos: -1, -3, -12, -9999). Para castigar.

### 8.3. Lista de mensajes

La **lista de mensajes** define las clases que podrán utilizarse en la lista de expresiones. Tiene el siguiente formato:

**CLASE;;PRIORIDAD;;NOMBRE;;PÁGINADEAVISO;;**

Donde cada parte representa lo siguiente:

1. «CLASE»: Indica el identificador de este tipo de edición. De este dato depende las clases de la lista de expresiones.
2. «PRIORIDAD»: Determina un orden entre las clases.
3. «NOMBRE»: El nombre de la clase.
4. «PÁGINADEAVISO»: La página que contiene el mensaje que será enviado.

Se permiten comentarios precedidos por el carácter «#».

## 8.4. Lista de exclusiones

La **lista de exclusiones** se encuentra por defecto en la subpágina «Exclusiones.css». En ella se van incorporando una a una (no es posible descartar varias páginas de una vez), aquellas que no deben ser vigiladas por el robot. Al igual que la lista de expresiones regulares, ésta es actualizada instantes siguientes a que sea modificada. Se permiten comentarios precedidos por el carácter «#».

## 8.5. Estadísticas

Una parte importante de los datos generados la constituyen las estadísticas.[90] En ellas se almacenan cada 2, 12 y 24 horas las cifras de «V» (vandalismos), «BL» (blanqueos), «P» (pruebas), «S» (contenido chocante), «B» (buenas), «M» (malas), «T» (total) y «D» (destruir). Mediante un *script* que capture estos datos históricos, es posible generar estadísticas como las mostradas en este documento.

## 8.6. Depuración

Para la correcta **depuración del programa**, se mantiene un *log* con cada acción que realiza el script. Esto incluye todas reversiones y las expresiones regulares que lo motivaron. Para facilitar su lectura, se almacena un log por cada día dentro del subdirectorio «botlogs/», ya que de lo contrario podría hacerse inmenso.

El siguiente es un extracto de un *log* real del 19 de enero de 2009:

```
Artículo: [[Pirámide de población]]
Fecha: 2009-01-19 00:50:41.979713
Puntuación: -5
Expresiones regulares:
p+e+[mn]+d+e+j+[ao]+s*
[haeiou]{6,}
p+i+n+c+h+e+s*
j+o+t+o+s*
```

El artículo afectado fue «Pirámide de población». Un vándalo introdujo una serie de palabras malsonantes que coincidieron con las expresiones regulares mostradas. La puntuación total fue de -5 puntos.

## 8.7. Para profundizar

Si se quiere profundizar en las posibilidades del programa se recomienda consultar la documentación del código realizada con Doxygen.



## 9. Manual de instalación

El *script* hace uso de dos bibliotecas o *frameworks* principales (y por supuesto del intérprete de Python, versión 2.4 o superior). Uno es pywikibot, una serie de funciones en Python para acceder y/o modificar los contenidos almacenados en un sitio web que use el software MediaWiki, en este caso Wikipedia en español. El otro es irclib, un conjunto de librerías para acceder a canales IRC.

Pywikibot se distribuye con una licencia MIT (disponible en la web <http://pywikibot.sourceforge.net/>), e irclib bajo licencia GPL (disponible en la web <http://python-irclib.sourceforge.net/>). Ambas licencias son ideales para su uso en proyectos de software libre, como en el caso que nos ocupa.

En primer lugar descargamos el intérprete de Python ideal para nuestro sistema operativo desde su web oficial <http://www.python.org/download/> y lo instalamos. Este paso no tiene dificultad alguna.

A continuación descargamos pywikibot e irclib a nuestro ordenador, debemos fusionarlos en un mismo directorio (irclib se ofrece en ficheros comprimidos listos para descargar, pero pywikibot debe ser bajado mediante un cliente SVN). También debemos incorporar en el mismo lugar los ficheros correspondientes a AVBOT (disponibles en <https://forja.rediris.es/projects/cusl3-avbot/> y <http://code.google.com/p/avbot/>). El siguiente paso es editar el fichero «user-config.py»<sup>25</sup> e incorporar las siguientes líneas:

```
mylang = 'NuestroIdioma'
usernames['Proyecto']['IdiomaDelProyecto'] = 'NombreDelBot'
```

Donde:

1. «NuestroIdioma» hace referencia al lenguaje en el que vamos a trabajar.
2. «Proyecto» contiene el proyecto e «IdiomaDelProyecto» el lenguaje de éste («NuestroIdioma» e «IdiomaDelProyecto» coinciden la mayoría de las veces).
3. «NombreDelBot» es el nombre de la cuenta de usuario que previamente debemos haber registrado en el sitio web y es un paso ineludible.

En el caso de AVBOT que va a vigilar los cambios que se producen en Wikipedia en español, el fichero quedaría de la siguiente forma:

```
mylang = 'es'
usernames['wikipedia']['es'] = 'AVBOT'
```

---

<sup>25</sup>Si se necesita más información acerca de este fichero y su configuración, se recomienda leer la documentación disponible en <http://meta.wikimedia.org/wiki/Pywikipedia>.

De esta manera sólo nos quedaría ejecutar el archivo `avbot.py` desde una terminal de MS-DOS o una shell de Linux. Un detalle importante es que la contraseña es necesaria para que el *bot* pueda editar las páginas, por ello, cuando ejecutemos el script, el *bot* hará una edición de prueba en la que se nos pedirá este dato.

Una vez introducida la contraseña, se mantendrá un archivo de *token* en un directorio especial de pywikipediabot, y no será necesario introducirla otra vez, a menos que borremos a mano dicho fichero o cambiemos la contraseña de la cuenta desde el sitio web.



## 10. Estadísticas

Desde mediados de marzo de 2008 hasta finales de ese mismo año AVBOT acumuló más de 5600 horas de trabajo. Al principio solamente estaba en funcionamiento durante el día, aunque más tarde fue instalado en un servidor en el que permanece 24/7 activo a menos que esté siendo depurado.

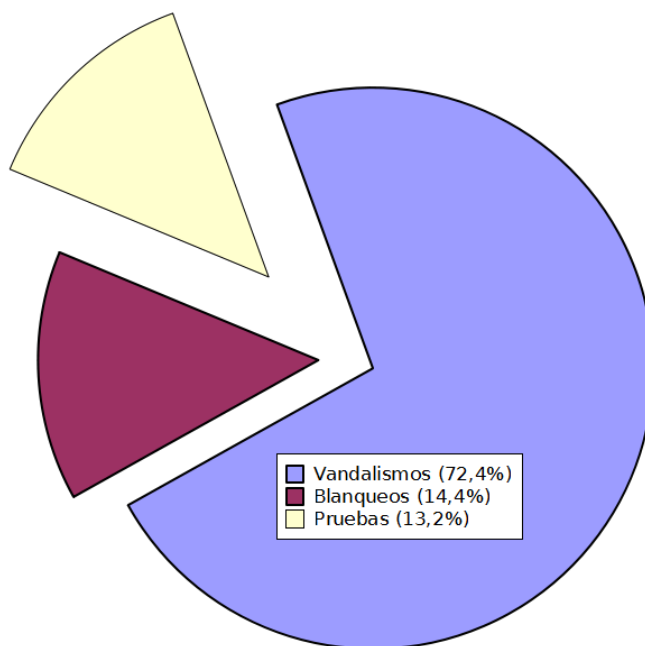


Figura 7: Motivos de las reversiones de AVBOT y su porcentaje sobre el total. Según datos de noviembre y diciembre de 2008, de las 10973 reversiones que realizó el script, 7950 (72,4 %) fueron provocadas por vandalismos, 1578 (14,4 %) por blanqueos y 1445 (13,2 %) debidas a pruebas de edición. El porcentaje de reversiones por pruebas podría haberse evitado si los usuarios hubiesen utilizado la zona habilitada para ello, llamada «Wikipedia:Zona de pruebas».

### 10.1. Artículos que han sufrido más intervenciones

La siguiente tabla recoge aquellos artículos en los que AVBOT ha intervenido un mayor número de veces durante el año 2008 por encontrar ediciones que debían ser deshechas según sus algoritmos.



#	Artículo	Intervenciones
1	Contaminación	76
2	Daddy Yankee	71
3	Computadora	59
	Baloncesto	59
	Roma	59
4	Aparato digestivo	54
5	Revolución industrial	50
6	Ensayo	46
7	Petróleo	45
8	Reciclaje	44
9	Droga	43
10	Habbo	40
11	Gran colisionador de hadrones	39
12	Renacimiento	38
	Naruto Uzumaki	38
	Microprocesador	38
	Hardware	38
13	Tecnología	37
	Música	37
	Energía	37
14	Wanda Nara	36
	Revolución francesa	36
	Pablo Neruda	36
	Ilustración	36
15	Leyes de Newton	35
	Fútbol	35
	Dragon Ball	35

Cuadro 1: Artículos con más intervenciones

Téngase en cuenta que en algunas intervenciones el robot ha podido eliminar varios vandalismos/blanqueos/pruebas en cascada.



## 11. ¿Qué es Wikipedia?

[Inicio de traducción]<sup>26</sup>

Wikipedia es un proyecto de la Fundación Wikimedia (una organización sin ánimo de lucro) para construir una enciclopedia libre<sup>27</sup> y políglota.<sup>28</sup> Su nombre es un portmanteau de las palabras *wiki* (una tecnología para la creación de sitios colaborativos) y *enciclopedia*. Los más de 11 millones de artículos de Wikipedia (2,6 millones en inglés)[17] han sido redactados colaborativamente por voluntarios de todo el mundo, y prácticamente todos pueden ser editados por cualquier persona que pueda acceder a Wikipedia.<sup>29</sup> Lanzada en enero de 2001 por Jimmy Wales y Larry Sanger,[?] es actualmente la mayor y más popular[18] obra de consulta en Internet.[20, 19]

Desde su concepción, Wikipedia no sólo ha ganado en popularidad sino que su éxito ha propiciado la aparición de proyectos hermanos. Existe, sin embargo, controversia sobre su fiabilidad.[21] En este sentido, la revista científica *Nature* declaró que la Wikipedia en inglés era casi tan exacta como la *Encyclopaedia Britannica* en artículos científicos.[22]

Actualmente Wikipedia presenta ediciones en 264 idiomas.[17] Veintitrés ediciones superan los 100.000 artículos: inglés, alemán, francés, polaco, japonés, italiano, holandés, portugués, español, ruso, sueco, chino, noruego, finés, catalán, ucraniano, turco, rumano, volapük, checo, húngaro, esperanto y eslovaco. La versión en alemán ha sido distribuida en DVD-ROM, y se tiene la intención de hacer una versión inglesa en DVD y en papel. Muchas de sus otras ediciones han sido replicadas a través de Internet (mediante «espejos») o han dado origen a enciclopedias derivadas (bifurcaciones) en otros sitios web.

Cuando la revista *Time* nombró a "You" (en español *Usted*) como "Persona del año" en 2006, reconociendo el éxito acelerado de la colaboración en línea y de la interacción entre millones de usuarios de todo el mundo, citó a Wikipedia como uno de los tres ejemplos de servicios "Web 2.0", junto con YouTube y MySpace.[23]

[Fin de traducción]

---

<sup>26</sup>Este primer apartado ha sido traducido y mejorado por el autor de esta memoria, a partir del artículo «Wikipedia» de Wikipedia en inglés (<http://en.wikipedia.org/w/index.php?oldid=258974889> con licencia GFDL) e incorporado a su homólogo en español (<http://es.wikipedia.org/w/index.php?diff=22723756&oldid=22702637>).

<sup>27</sup>Algunas versiones de Wikipedia (como la inglesa) integran contenido no libre (imágenes con *fair use*).[29]

<sup>28</sup>Jimmy Wales en una entrevista realizada por Slashdot en 2004 declaró lo siguiente: «*Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.*»[30] En español: «*Imagine un mundo en donde cada persona del planeta pueda tener acceso libre a la suma total de todo el conocimiento humano. Eso es lo que estamos haciendo.*»

<sup>29</sup>Wikipedia ha sido bloqueada en su totalidad o parcialmente en varias ocasiones y en diversas partes de mundo.[28, 26, 27]

## 11.1. Historia

Wikipedia se creó el 15 de enero de 2001,<sup>30</sup> como complemento de la enciclopedia escrita por expertos Nupedia. Pronto superó a ésta, dado que los artículos que se publicaban en Nupedia debían pasar un proceso de revisión por pares, mientras que en Wikipedia cualquier cambio se hacía visible al instante. La palabra Wikipedia se construyó a partir de los términos *wiki* (que proviene del hawaiano y significa rápido) y de *pedia* (que proviene del griego clásico, *paideia*, y significa educación).

La idea de reunir todo el saber humano no es nueva. La Biblioteca de Alejandría tenía similares objetivos, siglos más tarde, con la *Enciclopedia* de Denis Diderot volvería a repetirse. El desarrollo de Internet como herramienta de consulta mundial propició que el intercambio de información fuera posible a velocidades que antes jamás se habían concebido. Nupedia fue fundada por Jimmy Wales con la ayuda de un empleado, Larry Sanger, licenciado en filosofía. Era necesario el uso de una herramienta web que permitiese la edición de los artículos de una manera relativamente sencilla para un público amplio. Esto lo encontraron en el software WikiWikiWeb, programado por Ward Cunningham. Wikipedia tuvo un crecimiento muy rápido en sus comienzos. En su primer año de existencia se crearon 20.000 artículos, lo que hacía una media de 1500 artículos nuevos por mes. Este ratio ha ido fluctuando a lo largo del tiempo, pero hoy día es aun más acentuado para la versión inglesa que para el resto de idiomas.

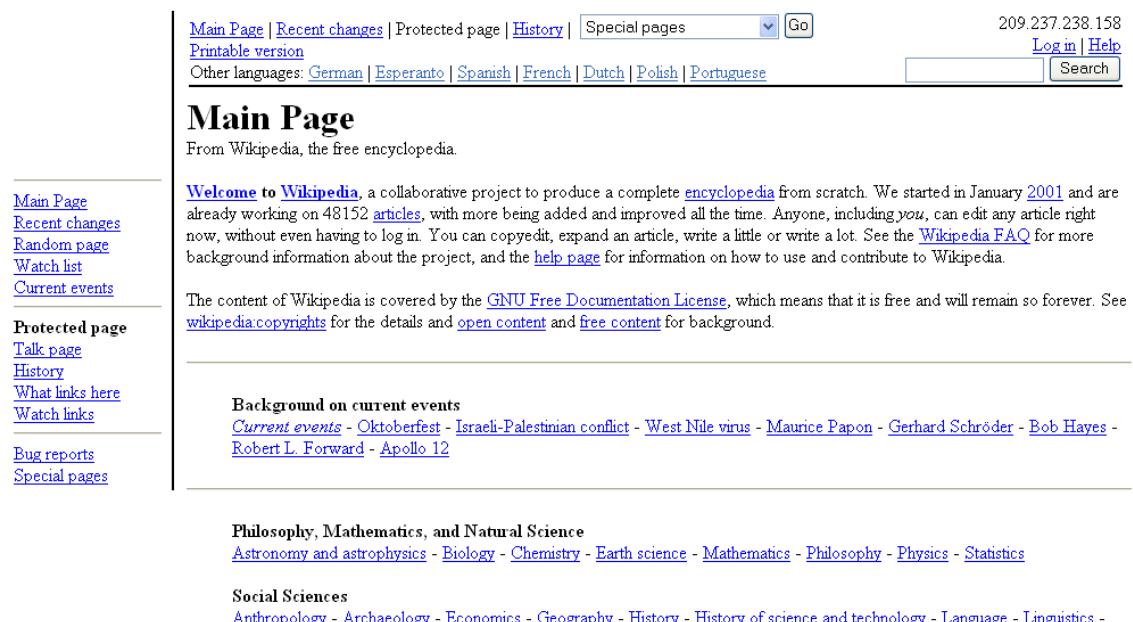


Figura 8: Página principal de Wikipedia en el año 2002

En abril de 2008, Wikipedia sobrepasó los 10 millones de artículos, incluyendo

<sup>30</sup>La primera en fundarse fue la versión en inglés. El 16 de marzo crearon la alemana y minutos después la catalana (aunque la germana no empezaría a tener artículos hasta abril).[7]

más de 2 millones en la versión en inglés. A día de hoy existen Wikipedias en más de 250 idiomas. A pesar de su permanente crecimiento, se han distribuido DVDs con contenidos de Wikipedia en alemán,[8] incluso se ha publicado alguna versión impresa con los mejores artículos.[9]

## 11.2. Comunidad

Los colaboradores de Wikipedia se llaman **wikipedistas**. La comunidad está formada por decenas de miles de voluntarios alrededor de todo el mundo. La edición con más usuarios registrados es la inglesa (8,6 millones), seguida por la española (923 mil), aunque solo un pequeño porcentaje de ellos está en activo.<sup>31</sup> Hay muchas maneras de colaborar con Wikipedia. A parte de la principal y evidente, crear artículos, también se pueden subir fotografías con licencia libre, corregir y ampliar textos que ya estén presentes, combatir el vandalismo, categorizar, *wikificar*,<sup>32</sup> entre muchas otras. Para ponerse en contacto con los bibliotecarios y solicitar sus funciones existen tableros especiales como el «Tablón de anuncios de los bibliotecarios» o «Vandalismo en curso» para denunciar actos de vandalismo reincidente. Para resolver problemas internos entre usuarios existen organismos específicos como el *Arbitration Committee* en la inglesa o el «Comité de Resolución de Conflictos» (CRC) en la española.

---

<sup>31</sup>Datos de enero de 2009.[24, 25]

<sup>32</sup>«*Wikificar*» es un neologismo. Significa dar formato, introducir enlaces, negritas, cursivas, poner categorías, a un artículo que carezca de ello.



Figura 9: Wikipedistas y miembros de Wikimedia Argentina  
 Autor: Ariel Kanterewicz. Licencia: Creative Commons Attribution-Share Alike 2.5 Argentina <http://creativecommons.org/licenses/by-sa/2.5/ar/>.

### 11.3. Control de calidad

En Wikipedia no existe ningún control editorial al estilo de las enciclopedias comunes. Esto no significa que no se dediquen todos los esfuerzos posibles a presentar información lo más fidedigna posible. Prueba de ello es la existencia de un extenso «Manual de estilo» que da orientaciones sobre cómo estructurar los artículos,[68] y de las clasificaciones de éstos en «esbozos»,[89] «artículos buenos»[88] y «artículos destacados»[87] según su grado de desarrollo.<sup>33</sup> Esto se hizo desde los primeros estadios de Wikipedia, resaltando los mejores logros de la comunidad.[85]

### 11.4. Críticas

En contraposición a las ventajas evidentes del proyecto, hay gente que opina que no todo es favorable en Wikipedia. Muchos piensan que dejar que cualquier persona edite la enciclopedia es peligroso, que podrían incorporar sin sentidos o material incorrecto. Esto es posible y además sucede continuamente, pero hay herramientas

<sup>33</sup>Sólo un pequeño porcentaje del total de artículos tienen la categoría de destacado. A enero de 2009 existían alrededor de 2300 de ellos. También existe una designación homóloga para las mejores imágenes libres disponibles en la enciclopedia, conocida como «featured pictures»[84, 83] y para sonidos[86], entre otras.

para combatirlo. Es fácil detectar ediciones puramente maliciosas para un humano. También es relativamente sencillo encontrar errores malintencionados, para ello existen páginas especiales dentro de Wikipedia como los «Cambios recientes», en los que se muestran las últimas ediciones que se han realizado. Además existen herramientas automáticas que son de gran ayuda, como la que se desarrolla en este PFC.

Hubo un estudio de la revista científica *Nature*,[22] en el que se comparaban los artículos de Wikipedia con sus homólogos en *Enciclopedia Britannica*. Los resultados arrojaban una calidad asombrosamente similar, aunque en ciertos pasajes *Britannica* superaba a Wikipedia, también se dio el caso de que Wikipedia era más completa en otros. *Britannica* exigió una rectificación, pues se negaban a aceptar los resultados del estudio, a lo que *Nature* se negó.

### 11.5. Proyectos hermanos

Wikipedia solo admite artículos enciclopédicos. Esto propició la aparición de nuevos proyectos con la misma filosofía, bajo el amparo de la Fundación Wikimedia, pero con otros contenidos. Tal es el caso de:

- Wikilibros: Libros de texto, manuales, tutoriales...
- Wikicitas: Frases célebres, refranes...
- Wikcionario: Diccionario libre.
- Wikisource: Obras literarias que ya han entrado en dominio público.
- Wikiespecies: Repertorio de especies biológicas.
- Wikinoticias: Fuente de noticias libres.
- Wikimedia Commons: Almacén de fotografías libres para el resto de proyectos.
- Meta-Wiki: Coordinación de los proyectos de la Fundación.
- Wikiversidad: Plataforma educativa libre.

### 11.6. Forks

Desde la aparición de Wikipedia han surgido numerosos proyectos similares, incluso alguno de ellos proviene de la escisión de algunos de sus voluntarios. Tal es el caso de *Enciclopedia Libre*, otro proyecto en español para la construcción de una enciclopedia libre. Se creó tras el rumor de que se iba a incorporar publicidad en Wikipedia. A día de hoy no se ha introducido publicidad en Wikipedia, ni se espera, así lo ha confirmado Jimbo Wales en numerosas ocasiones.

### 11.7. Wikipedia y los *bots*

Wikipedia es un sitio web muy grande en el que colaboran decenas de miles de personas cada día<sup>34</sup> y que requiere de un continuo mantenimiento. Diariamente se crean una gran cantidad de artículos nuevos (muchos de ellos deben ser marcados como prueba), se eliminan imágenes que no tienen licencia libre (y los artículos que estaban vinculados a esas imágenes deben ser desenlazados), se conectan artículos en distintos idiomas que hablan sobre un mismo tema a través de *interwikis*,<sup>[74]</sup> etc. Todas estas tareas son tediosas, y muchas de ellas serían interminables sino fuera por la ayuda de los *bots*.<sup>[72]</sup> A continuación se muestra una gráfica con la actividad en Wikipedia en español en 2008.

---

<sup>34</sup>Tan sólo en Wikipedia en español y durante el mes de diciembre de 2008, más de 13 mil usuarios realizaron alguna acción.<sup>[25]</sup>



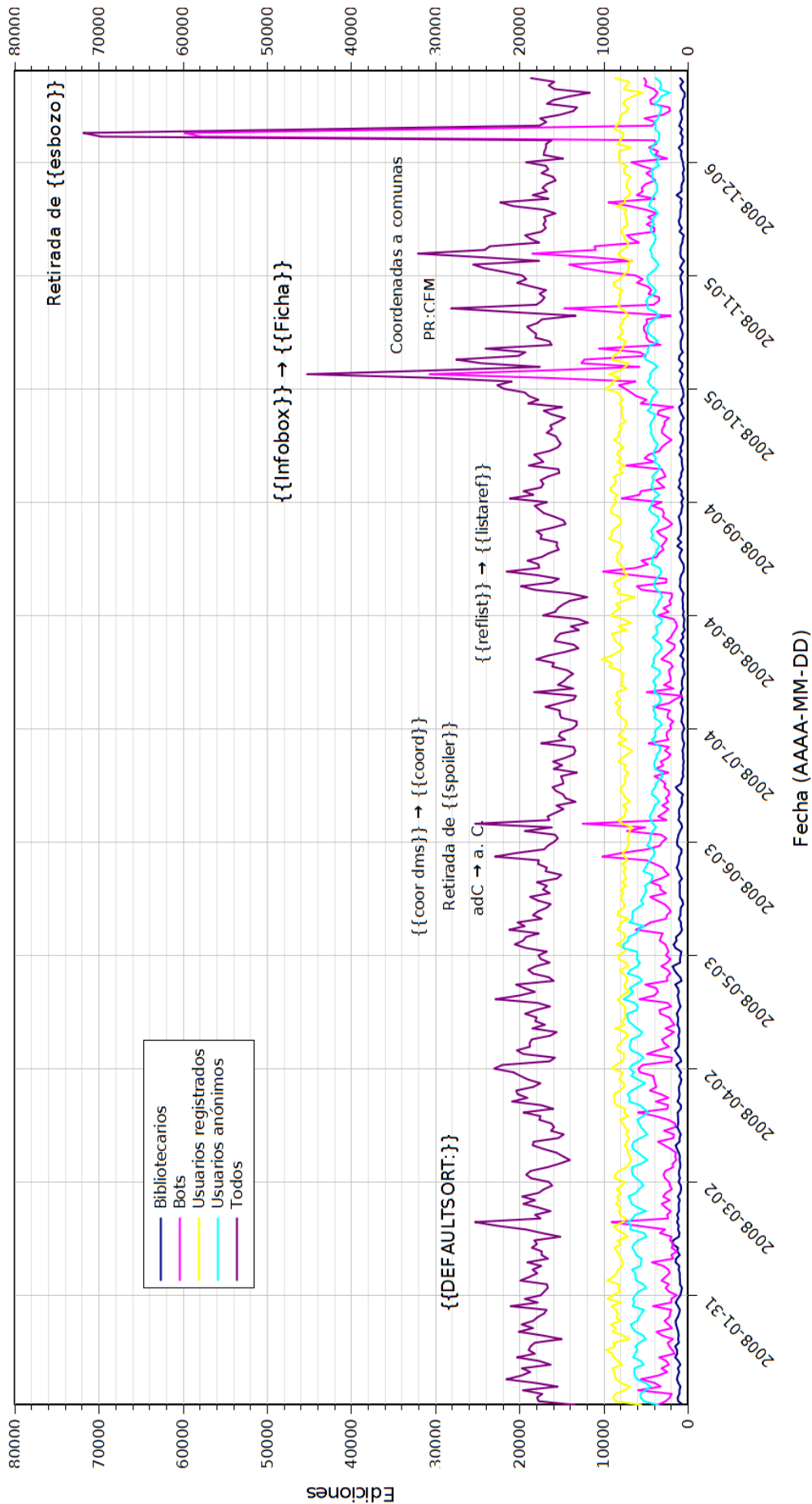


Figura 10: Ediciones a artículos por día durante 2008

La gráfica muestra la actividad durante el año 2008. Se han tomado las ediciones a artículos y anexos por ser las más representativas y relevantes, pues hacen mejorar los contenidos de la enciclopedia. Diariamente se realizan una media de 20 mil modificaciones tan sólo en Wikipedia en español (algunos miles más si se contasen los cambios en páginas de discusión). La línea rosa representa la actividad de los *bots*. Siempre hay cierto número de robots poniendo *interwikis* en los artículos (de ahí que siempre esté por encima de cero), y las subidas abruptas tienen lugar cuando se propone algún cambio masivo que afecta a una gran cantidad de artículos. Por ejemplo, en febrero se introdujo la instrucción «DEFAULTSORT» que permite organizar mejor las biografías según el orden «Apellido, Nombre», por ello hay un aumento de la actividad. Durante el año se produjeron otra serie de cambios masivos, siempre con la ayuda de *bots*, pero el más destacado se realizó en diciembre, cuando la comunidad decidió que no deseaba hacer una distinción de artículos según su grado de desarrollo y se retiró la plantilla «esbozo,» produciéndose una media de más de 70 mil ediciones diarias.



## 12. Curso de edición de páginas MediaWiki

Para comprender mejor el funcionamiento y el contexto del proyecto, se incluye un curso de edición de sitios MediaWiki.

### 12.1. Edición básica

Para editar un artículo basta con pulsar la pestaña «editar» situada en la parte superior de la pantalla. Es posible que un artículo esté protegido o semiprotegido,<sup>35</sup> en ese caso sólo podrían editarlo los administradores o los usuarios registrados, respectivamente.

#### 12.1.1. Formatos de texto

Es posible dar formato al texto de los artículos.

Código	Salida
<code>''Negritas''</code>	<b>Negritas</b>
<code>''Cursivas''</code>	<i>Cursivas</i>
<code>'''Cursivas en negrita'''</code>	<b><i>Cursivas en negrita</i></b>

Cuadro 2: Formato de texto

Algunas recomendaciones:

- No abusar de **negritas** y *cursivas* pues **suelen *dificultar*** la lectura *del texto*.

#### 12.1.2. Enlaces a otros artículos de Wikipedia

Podemos enlazar con otros artículos como si de una página web se tratase.

Código
<code>Da Vinci nació en [[Pisa]]</code>
<code>Da Vinci era [[Renacimiento renacentista]]</code>
<code>Leonardo estudió muchas [[ciencia]]s</code>

Cuadro 3: Enlaces

---

<sup>35</sup>Los artículos se protegen contra la edición por diversos motivos: vandalismo continuado, diferencias sin consensuar sobre el contenido, alto riesgo de vandalismo, etc. Ejemplos de páginas frecuentemente protegidas contra la edición son «Cataluña», «Augusto Pinochet» e «Idioma valenciano», entre muchas otras. La página principal de Wikipedia, la portada, siempre está protegida y sólo pueden editarla los bibliotecarios.

Algunas recomendaciones:

- No conviene llenar todo el texto de enlaces. Enlazar solo aquello que aporte información al artículo.
- No hace falta poner «\_» para separar palabras. «Leonardo da Vinci» es correcto, «Leonardo\_da\_Vinci» es incorrecto aunque funcione.
- Debemos ser fieles a mayúsculas y minúsculas. Si queremos enlazar con «Leonardo da Vinci» no debemos poner «leonardo da vinci», pues lo más probable es que no funcione (más información sobre esto en el apartado sobre redirecciones).

### 12.1.3. Enlaces a otros artículos de Wikipedia en otros idiomas u otros proyectos de Wikimedia

Es posible enlazar con otros idiomas u otros proyectos como Wikiquote, Wiktionary, etc.

Código	Comentario
<code>[[fr:Biologie]]</code>	Enlace a Wikipedia en francés
<code>[[q:de:Albert Einstein]]</code>	Wikiquote en alemán
<code>[[wikt:vi:nhà]]</code>	Wiktionary en vietnamita

Cuadro 4: Enlaces a proyectos hermanos

### 12.1.4. Enlaces a páginas web externas a Wikipedia

Los llamados «enlaces externos» son vínculos que apuntan a páginas web de Internet que permiten ampliar la información disponible en los artículos.

Código	Salida
<code>[http://www.ejemplo.com Ejemplo S.A.]</code>	Ejemplo S.A.
<code>[http://www.ejemplo.com]</code>	[1]
<code>http://www.ejemplo.com</code>	http://www.ejemplo.com

Cuadro 5: Enlaces externos

### 12.1.5. Listas

Hay dos tipos principales de listas.

Código	Salida	Comentarios
# Elemento uno # Elemento dos # Elemento tres	1. Elemento uno 2. Elemento dos 3. Elemento tres	Lista numerada
* Elemento uno * Elemento dos * Elemento tres	· Elemento uno · Elemento dos · Elemento tres	

Cuadro 6: Listas

### 12.1.6. Secciones

Es posible y necesario estructurar el texto en secciones y subsecciones.

Código	Salida	Comentarios
= Sección 1 = == Sección 2 == === Sección 3 ===	<b>Sección 1</b> Sección 2 Sección 3	Es posible llegar hasta el sexto nivel

Cuadro 7: Secciones

Algunas recomendaciones:

- No usar el tamaño 1 ni tamaños a partir del 5. El 1 es demasiado grande y a partir del 5 la estructura de secciones se hace incómoda.

## 12.2. Imágenes

Antes de incluir una imagen en Wikipedia debes cerciorarte de que tiene una licencia libre compatible con los proyectos de Wikimedia. Las principales licencias aceptadas son: Dominio Público, GFDL, CC-by, CC-by-sa. Si cumple este requisito y además la imagen es útil para Wikipedia (no debes subir imágenes de tus vacaciones con la familia o de las fiestas con tus amigos, para eso existen otras webs y *fotologs*), sigue los siguientes pasos:

1. Subir la imagen al *hosting* propio de la Fundación Wikimedia, <http://commons.wikimedia.org>, más concretamente el formulario de subida <http://commons.wikimedia.org/wiki/Special:Upload>
2. Ir al artículo en cuestión e incluir una de las siguientes posibilidades:
  - a) Para imágenes a la derecha:  
`[[Archivo:TorreDeSegovia.jpg|thumb|right|200px|Torre de Segovia]]`
  - b) Para imágenes a la izquierda:  
`[[Archivo:TorreDeSegovia.jpg|thumb|left|200px|Torre de Segovia]]`

Otras posibilidades:

- Usar `<gallery></gallery>`. Permite insertar una galería de imágenes, todas con el mismo tamaño y alineadas. Ejemplo:  
`<gallery>`  
`Archivo:Imagen1.jpg|Esta es la imagen 1`  
`Archivo:Imagen2.jpg|Esta es la imagen 2`  
`Archivo:Imagen3.jpg|Esta es la imagen 3`  
`</gallery>`

Algunas recomendaciones:

- Se suelen poner las imágenes siempre a la derecha.
- No se deben sobrepasar los 200px por imagen a menos que sea necesario (por ejemplo en imágenes panorámicas, obras de arte de gran ancho), de lo contrario el artículo puede quedar ilegible en resoluciones de pantalla pequeñas (800x600).
- No colocar muchas imágenes en un mismo artículo. Mantener un equilibrio.

### 12.3. Categorías

Es posible y necesario categorizar los artículos.

Código	Comentarios
<code>[[Category:Municipios de Cádiz]]</code>	Se categoriza en "Municipios de Cádiz".

Cuadro 8: Categorías

Alguna recomendaciones:

1. No categorizar de manera redundante. No incluir a Algeciras en "Municipios de Cádiz" y en "Municipios de España" a la vez. Usar la más descriptiva.

2. El código de las categorías se pone al final del artículo, justo antes de los *interwikis*.

## 12.4. *Interwikis*

Podemos enlazar un artículo con su homólogo en otra Wikipedia en otro idioma.

Código	Comentarios
<code>[[en:Spain]]</code>	El artículo «España» quedará enlazado con su homólogo en inglés.
<code>[[ru:Испания]]</code>	El artículo «España» quedará enlazado con su homólogo en ruso.

Cuadro 9: *Interwikis*

Alguna recomendaciones:

- No tenemos que adivinar si existe tal artículo en tal idioma. Basta con ir al artículo en Wikipedia en inglés y copiar los *interwikis* desde allí.
- Los *interwikis* van al final del artículo, después de las categorías.

## 12.5. Tablas

MediaWiki soporta la creación de tablas en formato HTML, aunque conviene utilizar su propia sintaxis.

Toda tabla comienza con «{|» y termina con «|}». Justo detrás de «{|» es posible insertar estilo CSS. Cada fila se genera poniendo «|-» y la separación entre celdas de una misma fila se realiza con «||». Un ejemplo:

```
{| style=estilo CSS
! CabeceraA !! CabeceraB !! CabeceraC
|-
| Celda1Fila1 || Celda2Fila1 || Celda3Fila1
|-
| Celda1Fila2 || Celda2Fila2 || Celda3Fila2
|}
```

Resultado:

CabeceraA	CabeceraB	CabeceraC
Celda1Fila1	Celda2Fila1	Celda3Fila1
Celda1Fila2	Celda2Fila2	Celda3Fila2

Es posible utilizar las propiedades «colspan», «rowspan», «align» y «valign». Para ello basta con colocar el código antes de la celda que queremos sea afectada.

Ejemplo con «colspan»:

```
{| style=estilo CSS
! CabeceraA !! CabeceraB !! CabeceraC
|-
| colspan=2 | Celda1Fila1 || Celda2Fila1
|-
| Celda1Fila2 || Celda2Fila2 || Celda3Fila2
|}
```

Resultado:

CabeceraA	CabeceraB	CabeceraC
Celda1Fila1		Celda2Fila1
Celda1Fila2	Celda2Fila2	Celda3Fila2

## 12.6. Plantillas

Las plantillas son páginas que pueden ser incluidas en otras páginas. Su cometido es evitar que un mismo texto o código tenga que ser repetido en infinidad de lugares, lo que podría provocar errores de consistencia. Además, sus funciones se pueden flexibilizar mediante parámetros.

Toda plantilla comienza con «`{{`» seguido del nombre y finaliza con «`}}`». Debemos tener en cuenta que una plantilla puede contener cualquier cosa en su interior: imágenes, texto, otras plantillas... Veamos un ejemplo, supongamos que el contenido de la página «Plantilla:MiPlantilla» es el siguiente:

Hola `{{{1}}}`, ¿cómo estás? Yo `{{{2}}}`.

Ahora, si fuéramos a cualquier otra página del sitio, y pusieramos:

`{{MiPlantilla|Pepe|bien}}`

El resultado sería:

Hola Pepe, ¿cómo estás? Yo bien.

Podríamos hacer que si no se le pasa el «parámetro 2», escribiera un texto por defecto. El código de la plantilla debería ser:

Hola `{{{1}}}`, ¿cómo estás? Yo `{{{2|mal}}}`.



De este modo, si llamamos a la plantilla:

```
{{MiPlantilla|Juan}}
```

Veríamos:

```
Hola Juan, ¿cómo estas? Yo mal.
```

Este es un ejemplo muy simple y carente de utilidad. Las plantillas pueden llegar a ser tan complejas como sea necesario y en Wikipedia se hace un uso intensivo de ellas, sobre todo para poner tablas de información en los artículos llamadas «Infobox» o «Fichas».

Si quisieramos que una parte de la plantilla no fuera visible en la página en la que se integra, bastaría con poner ese fragmento entre las etiquetas `<noinclude>` y `</noinclude>`.

## 12.7. Palabras mágicas

El software MediaWiki proporciona una serie de instrucciones llamadas «palabras mágicas» con las que es posible realizar cálculos con las *parserFunctions*. Son usadas ampliamente en las plantillas.

## 12.8. Para profundizar

1. Ayuda de Wikipedia en inglés <http://en.wikipedia.org/wiki/Help:Contents>
2. Preguntas frecuentes [http://en.wikipedia.org/wiki/Wikipedia:FAQ\\_Index](http://en.wikipedia.org/wiki/Wikipedia:FAQ_Index)
3. Ayuda para tablas <http://es.wikipedia.org/wiki/Ayuda:Tablas>
4. Ayuda para plantillas <http://es.wikipedia.org/wiki/Ayuda:Plantillas>

12.7.1.    Tiempo

Código	Salida	Comentarios
{{CURRENTDAY}}	1	Día del mes
{{CURRENTDAY2}}	01	Día del mes pero rellenando con ceros los días menores de 10. Ej: 01, 02, ...
{{CURRENTDAYNAME}}	jueves	Nombre del día de la semana
{{CURRENTDOW}}	4	Número del día de la semana
{{CURRENTMONTH}}	01	Número del mes
{{CURRENTMONTHABBREV}}	ene	Abreviatura del nombre del mes
{{CURRENTMONTHNAME}}	enero	Nombre del mes
{{CURRENTTIME}}	00:00	Hora en formato 24h
{{CURRENTHOUR}}	00	Hora actual
{{CURRENTWEEK}}	1	Semana del año
{{CURRENTYEAR}}	2009	Año
{{CURRENTTIMESTAMP}}	20090101000000	Timestamp UNIX

Cuadro 10: Palabras mágicas referentes a unidades temporales

12.7.2. Estadísticas

Código	Salida	Comentarios
{{CURRENTVERSION}}	1.12.0	Versión del software MediaWiki
{{NUMBEROFEDITS}}	314.159	Número de ediciones realizadas
{{NUMBEROFEDITS:R}}	314159	Igual que el anterior pero <i>Machine-Readable</i> <sup>a</sup>
{{NUMBEROFARTICLES}}	4.023	Número de artículos (páginas en el espacio de nombres 0)
{{NUMBEROFARTICLES:R}}	4023	Igual que el anterior pero <i>Machine-Readable</i>
{{NUMBEROFPAGES}}	6.341	Número total de páginas (incluye artículos, páginas de discusión, etc)
{{NUMBEROFPAGES:R}}	6341	Igual que el anterior pero <i>Machine-Readable</i>
{{NUMBEROFFILES}}	666	Número de ficheros subidos
{{NUMBEROFFILES:R}}	666	Igual que el anterior pero <i>Machine-Readable</i>
{{NUMBEROFUSERS}}	42	Número de cuentas de usuario registradas
{{NUMBEROFUSERS:R}}	42	Igual que el anterior pero <i>Machine-Readable</i>
{{NUMBEROFADMINS}}	8	Número de usuarios con privilegios de administrador
{{NUMBEROFADMINS:R}}	8	Igual que el anterior pero <i>Machine-Readable</i>

Cuadro 11: Palabras mágicas referentes a datos estadísticos

<sup>a</sup>Las versiones *Machine-Readable* de estas palabras mágicas se utilizan para realizar cálculos con la extensión *ParserFunctions*.<sup>[31]</sup>



## 13. GNU Free Documentation License

Extraído de: <http://www.gnu.org/licenses/fdl-1.3.txt>

GNU Free Documentation License Version 1.3, 3 November 2008

Copyright (C) 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

<<http://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

### PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

### 13.1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, L<sup>A</sup>T<sub>E</sub>X input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page

as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## **13.2. VERBATIM COPYING**

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## **13.3. COPYING IN QUANTITY**

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they

preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## 13.4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. C. State on the Title page the name of the publisher of the Modified Version, as the publisher. D. Preserve all the copyright notices of the Document. E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. F. Include, immediately after the copyright notices, a license notice giving the public permission to use



the Modified Version under the terms of this License, in the form shown in the Addendum below. G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. H. Include an unaltered copy of this License. I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the

previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

### **13.5. COMBINING DOCUMENTS**

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

### **13.6. COLLECTIONS OF DOCUMENTS**

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

### **13.7. AGGREGATION WITH INDEPENDENT WORKS**

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium,

is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

### **13.8. TRANSLATION**

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

### **13.9. TERMINATION**

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

### **13.10. FUTURE REVISIONS OF THIS LICENSE**

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.